Escola de Camins
Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports
UPC BARCELONATECH

Automated driving and the Big Five Personality Traits: One size fits none.

TREBALL FINAL DE GRAU

Treball realitzat per:
Alba Rodríguez Sayrol

Dirigit per:
Daniel Rodríguez Aranda

Grau en:
Enginyeria Civil

Barcelona, 30 de setembre de 2019.

DECA - Departament d'Enginyeria Civil i Ambiental

Acknowledgements

I would first like to thank my tutor from UPC Daniel Rodríguez Aranda and my supervisor from TU Delft Daniël Heijkoop.

In addition, I would like to thank my parents and my friends, who are always the best support.

Abstract

This thesis consists on finding the selection of the most representative profiles of drivers for studying how different personalities behave with the new technologies of automated driving. This work was carried out under the supervision of professor Daniël Heikoop at TU Delft with the main goal to investigate the behaviour of drivers based on the Big Five Personalities classification, Big Five for short. This is a categorization based on the composition of five personality dimensions, which were defined independently by psychology researchers during the 2nd half of the 20th Century. Participants of the project, recruited according to their Big Five results and believed to be representative of drivers in the Netherlands, will carry on the experiment, which will be performed in the driving simulator located at TU Delft.

The work first begins with a description of the state of the art of the theory of the Big Five Personality traits, as well as different approaches of personality classification such as Freud's theories, Eysenck's theory and Cattell's 16PF. The thesis continues to describe the development of a model for participant selection. Using this model, contributors are recruited among a variety of stakeholder employees and/or participant pools. Their personality is analyzed through a questionnaire. Then, the most suitable subjects to take part in the experiment are selected by creating classification lists and choosing the top 10 representatives of each personality groups. Two possible implementations are proposed.

The project tries to find a new approach to select candidates based on a mathematical analysis of the data set using the concept of Principal Component Analysis, an alternative to the Big Five's. Finally, both methods are compared.

Alba Rodríguez Sayrol

Barcelona, September 2019

Resum

Aquest projecte consisteix en trobar la selecció dels perfils més representatius de conductors per estudiar com es comporten diferents personalitats amb la conducció automatitzada. El projecte s'ha dut a terme sota la supervisió del professor Daniël Heikoop de la TU Delft amb l'objectiu principal d'investigar el comportament dels conductors basat en la classificació del Model dels cinc factors, Els Cinc Grans per a resumir. Es tracta d'una categorització basada en la composició de cinc dimensions de la personalitat, definides per investigadors de psicologia durant la segona meitat del segle XX. Els participants del projecte, seleccionats en base als seus resultats del Model dels cinc factors i representatius dels conductors als Països Baixos, continuaran l'experiment, que es realitzarà un simulador de conducció situat a TU Delft.

En primer lloc, el projecte comença amb una descripció de l'estat de l'art de les teories de personalitat més representatives del moment. Començant per el Model dels cinc factors i seguit per diferents enfocaments de classificació de la personalitat, com ara les teories de Freud, la teoria d'Eysenck i el 16PF de Cattell. A continuació, descriu el desenvolupament d'un model de selecció de participants. Amb aquest model, els col·laboradors són seleccionats entre diversos empleats interessats i / o agrupacions de participants. La seva personalitat s'analitza mitjançant un qüestionari. Més endavant, es seleccionen els més adequats per participar a l'experiment. La selecció es fa mitjançant la creació de llistes de classificació, on els 10 primers candidats seran escollits com a representants de cada grup de personalitat. Es proposen dues possibles implementacions.

El projecte intenta trobar un nou enfocament per seleccionar els candidats a partir d'una anàlisi matemàtica del conjunt de dades mitjançant el concepte d'Anàlisi de Components Principals, una alternativa al Model dels cinc factors. Finalment, es comparen els dos mètodes.

Contents

Abstract i

Resum iii

List of Figures vii

List of Tables ix

1 Introduction 1

2 Autonomous Vehicles 3

2.1 What is an automated vehicle . . . 3

2.2 Why automated vehicles are the future . . . 4

2.3 How an automated vehicle works . . . 6

2.3.1 General Insight. . . . 6

2.3.2 Communication systems. . . . 8

2.4 Relevance of the AV in the project . . . 10

3 Personality theories and driving behavior 11

3.1 Personality theories . . . 11

3.2 Big Five personality traits . . . 12

3.2.1 History . . . 13

3.2.2 Descriptions of the particular personality traits . . . 13

3.3 Other theories. . . . 14

3.3.1 Freud's Theories . . . 14

3.3.2 Eysenck's Theory . . . 15

3.3.3 Catell's 16 Personality Factors . . . 17

3.4 Correlation between personality and driving . . . 18

3.4.1 Behaviour in traffic. . . . 18

3.4.2 Personality and driving . . . 21

4 Selection based on the Big Five Scoring **23**

4.1 Participant selection . . . 23

4.1.1 Making the questionnaire . . . 23

4.1.2 Selecting the candidates . . . 24

4.2 Results . . . 25

5 Selection based on Principal Component Analysis **31**

5.1 Introduction to the PCA . . . 31

5.2 PCA . . . 33

5.2.1 General Insight . . . 33

5.2.2 Mathematical description . . . 34

5.3 PCA Applied to the Big Five Inventory . . . 37

5.3.1 Results of the PCA . . . 37

5.3.2 Alternative lists of candidates . . . 41

6 Conclusions and Future work/Trends **43**

Bibliography **45**

A Automated driving and the Big Five Personality Traits in English **49**

B Automated driving and the Big Five Personality Traits in Dutch **61**

C Big Five Inventory in English **73**

D Big Five Inventory in Dutch **77**

E Big Five Invenroty in Spanish **81**

F Selection algorithm 1 for Tu Delft selection, based on the Big Five Inventory **91**

G Selection algorithm 2 for Tu Delft selection, based on the Big Five Inventory **95**

H Selection algorithm of the PCA **97**

List of Figures

2.1 Different levels of automation [1]. . . . 4
2.2 Reasons why automated vehicles are the future [1]. . . . 4
2.3 Number of fatalities between 1960 to 2016 in Spain. Data from the DGT [3]. . . . 5
2.4 Robocabs. Self-driving service [1]. . . . 6
2.5 Connectivity between vehicles [4]. . . . 8
2.6 Existing alternatives to wireless technologies [6]. . . . 9
2.7 Inter-vehicle connectivity [6]. . . . 9

3.1 The Big Five groups and its characteristics [14]. . . . 12
3.2 Representation of the three aspects of the mind [15]. . . . 15
3.3 Different characteristics of both dimensions [19]. . . . 16
3.4 The 16 Personality types [21]. . . . 17
3.5 Model of mental workload, task performance and demand [23]. . . . 19
3.6 Variables of task demands and applied capability [24] 20
3.7 Work situations described by the tasks, task demands and capabilities [9]. . . . 20

4.1 Answers distributions of the candidates. O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness and N:Neuroticism. . . . 25
4.2 Histogram for category Openness. . . . 26
4.3 Histogram for category Conscientiousness. . . . 27
4.4 Histogram for category Extraversion. . . . 27
4.5 Histogram for category Agreeableness. . . . 28
4.6 Histogram for category Neuroticism. . . . 28
4.7 Distribution of the candidates with the Selection algorithm 2. . . . 29
4.8 Lists of the candidates using the two selection algorithms. . . . 30

5.1 PCA example [31]. . . . 32
5.2 Example of a negative correlated cells. . . . 33
5.3 Example of a PCA plot. . . . 34

5.4 Eigenvectors of the sample. 35
5.5 Principal Components that give good information. 37
5.6 Principal components with less accuracy. 37
5.7 Variance of the 44 questions. 38
5.8 Plot of the PC1 and PC2. 39
5.9 Distribution of the eigenvalues. 40
5.10 Plot of the highest three eigenvalues. 40
5.11 Lists with the Big Five Selection 2 and PCA. 41

List of Tables

2.1 Different types of sensors [1]. 7

4.1 Mean and Standard Deviation of categories. 26

5.1 Transcription of the gens in two cells. 33

1

Introduction

The subject matter of automated vehicles is becoming predominant among the topics of urban mobility. Leading cities are striving to become pioneers in the field, as well as automobile companies. The advantages that this new technology offers are many, from security to environmental perspective. But still there is a long way to go before they become a reality.

The main goal of this project is to delve deeper into the issue of Automated Vehicles. In order to gain this knowledge, TU Delft is preparing a project to investigate the behavior of drivers of automated vehicles, based on the Big Five Personalities. The different selected candidates are going to participate in a driving simulator experiment, to test their driving behavior.

The project is divided in four parts. First of all, a general insight of the automated vehicles is given. This part describes what is understood by automated vehicles and the different levels of automation that it comprises, the reasons why automated vehicles are envisioned for the future, how this technology works and the relevance of the automated vehicles in the project.

Secondly, the project provides a general view of the different personality theories currently studied by psychologists with special emphasis in the Big Five personality traits, the one used to select the candidates. This chapter introduces personality features and their importance. Moreover, it comprises some clarifications regarding the behavior of drivers in traffic and the correlation between personality and driving.

The third part of this project is focused on the selection process for the TU Delft investigation. There are two questionnaires, one in English and one in Dutch designed to quantify the participant's personality,

allowing the selection of the most relevant profiles. The questionnaire includes demographic information and the Big Five inventory. The selection method is implemented Matlab, a general purpose data processing platform. The developed code provides two sets of lists following different criteria.

Next, the project proposes an alternative method of selection based on Principal Component Analysis which performs the classification based on the statistical analysis of the questions. Principal Component Analysis is a method of dimensionality reduction that takes into account the correlation of the data. Matlab is also used in order to study the Big Five Inventory, in this case, to find the principal components, or projections, that will provide the maximum information. Once the principal components are found, it is possible to make new lists with the most relevant candidates.

Finally, the project compares the two classifications, the one made using the big five standard classifications and the second one, which uses the principal components of the data extracted by mathematical analysis and some conclusions are drawn.

2

Autonomous Vehicles

2.1. What is an automated vehicle

An autonomous vehicle (AV) is a vehicle capable of imitating and improving human car driving and control capabilities. It possesses different modules including AI (Artificial Intelligence) techniques that are able to perceive the environment around and navigate through it.

There are many levels of automation possible in an AV, its highest level is able to bring you to a chosen destination without any mechanical operation of the vehicle. The perception of the environment is made though a complex technology such as laser, radar, LIDAR, global positioning system and computer vision. The different levels of automation in a vehicle are [1]:

Level 0: No automation.

Level 1: Driver Assistance. The driver is in charge of single control functions such as speed selection, braking or lane keeping .

Level 2: Partial Automation. More than one control function is automated. An AI is added in order to improve the safety and convenience. The driver is still responsible for the car but the platform can perform automated manoeuvres [2].

Level 3: Conditional Automation. The driver is expected to respond to requests to take control. The AI of the vehicle can perform the dynamic driving task in certain modes without supervision.

Level 4: High automation. The driver is not expected to respond to requests to take control. The vehicle

can perform the dynamic driving task in certain modes .

Level 5: Full automation. No driver needed. The vehicle can perform the dynamic driving task in all modes that can also be managed by a human driver.

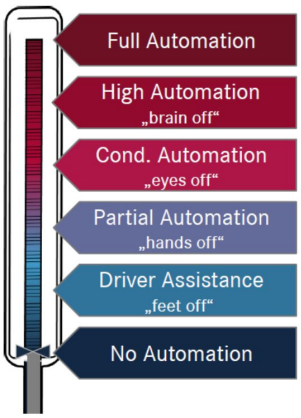


Figure 2.1: Different levels of automation [1].

2.2. Why automated vehicles are the future

The reasons behind the development of automated vehicles are many, but the principal are [1]:

Figure 2.2: Reasons why automated vehicles are the future [1].

- **The safety case:** automation to improve traffic safety.

  Safety is and will always be the main priority on the road. A graph from the *Dirección General de Tráfico* (DGT), an autonomous body of the government of Spain which is responsible for the road policy and

traffic control, shows the decrease in the number of fatalities since 1990. Even though it is improving, the goal is to avoid all casualties.

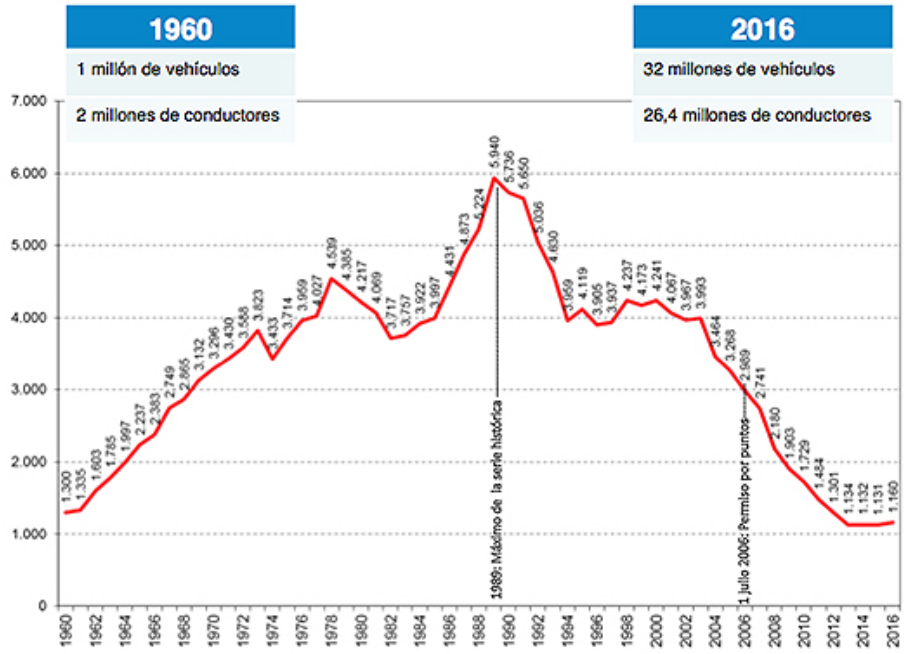


Figure 2.3: Number of fatalities between 1960 to 2016 in Spain. Data from the DGT [3].

Currently, almost all accidents are due to human error. In order to reduce or, in the best of the cases, avoid them, automatic systems should take control. There are two focus in the automation challenge. In the one hand, these systems should avoid the errors that humans make. On the other hand, they should not make mistakes where humans react properly.

- **The mobility case:** mobility for those who are not able to drive.

  In the highest level of automation, where there is an automated transport system under operator control, people with no ability or no permission to drive can benefit from being transported at a reasonable cost. This cloud apply to disabled people, seniors, juniors or people without driving license. This is also a safer option for people who drive under the influence of drugs, which is one of the main causes for road accidents.

- **The comfort case:** automation to meet customer demands.

  The private vehicle is usually the most enjoyable and comfortable way to travel short and medium distances. The reality today is that driving is impaired by dense congestion, slow traffic and long travel times. In crowded cities, mass transportation is the best way to reach a destination rapidly. Individual transportation may be slower, but provides the benefit of privacy. Time spent in the transport has

become a part of the day, vehicle automation would let people spend their transport time in other tasks and create a comfortable experience.

- **The disruption case**: forms of on-demand mobility.

  Automated vehicles, also known as Robocabs, which are self-driving cars, will eventually cover the benefits of car sharing and taxis. Robocabs will alleviate the problems that these groups face like limited coverage, need for parking, cost and the need of a driver. The benefits of sharing automated vehicles are many. First of all, the user convenience, the option to choose the right type of car where needed, when needed, as long as needed. Secondly, the existence of these cars will lead to a fewer cars parked, which means fewer cars in total. Moreover, the lower cost associated to sharing the cost and maintenance of a vehicle creates efficiency gains. Finally, the improved use of the mobility infrastructure and new urbanization models will decrease the space consumption.

Figure 2.4: Robocabs. Self-driving service [1].

2.3. How an automated vehicle works

2.3.1. General Insight

There are many systems that work in conjunction with each other to control an automated vehicle. Google, Uber, Tesla, Nissan, and other major automakers, researchers, and technology companies have developed numerous self-driving technologies. While design details differ, most automated vehicles create and internal map of their surroundings, based on a large array of sensors, like Radar, Lidar and Sonar. AV systems are also composed of actuators, a brain, and a back-end [1].

- **Sensors.** Autonomous navigation requires a wide-ranging vision and interpretation of the surroundings. Many sensors are required to guarantee a secure navigation. They can be classified in three groups:

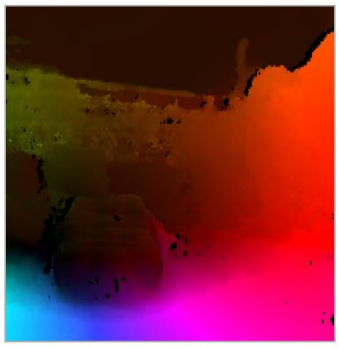
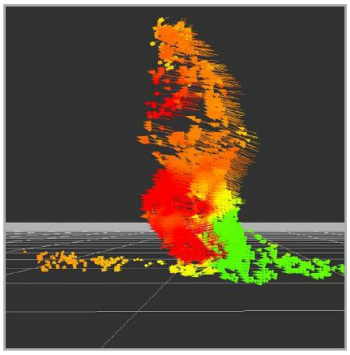
Dense Spacial Data	Classification	Motion Analysis
		

Table 2.1: Different types of sensors [1].

Dense Spatial data sensors which provide a detailed description of the surroundings under all possible weather and visibility conditions. There are many technology issues regarding this type of sensors due to the very large and different environmental conditions in which they need to operate.

Computer vision sensors for Classification which provide semantic labeling of objects, mostly vehicles, infrastructure, animals and pedestrians that are encountered during the driving.

Computer vision sensors for Motion Analysis which provide movement calculation and prediction of the moving objects recognized. Some of the issues of the motion analysis are related to the prediction of road user behavior.

- **Vehicle Control.** A control unit at the vehicle mechanizes all actions, deciding which options should be taken depending on the specific situation. The decisions are based on all the data and information available. The control unit is composed of four different entities:

  Sensor fusion and location process that provides data regarding vehicle position.

  Situation analysis process in charge of prediction of likely events around the vehicle.

  Motion planning process in charge of deciding proper planning of vehicle trajectory.

  Vehicle administer process in charge of interfacing to the vehicle actuators.

- **Redundant Actuators.** All driving actions need to be controlled automatically since AV must provide a safe transport for the users inside and outside the vehicle. In order to prevent any failures or errors, AV incorporate redundant actuators. Examples are redundant power and communication units, with two independent power sources and disjoint communication links. Moreover, redundant drive state sensors are required, with replicated sensors for vehicle movement detection. Also, AV have a fail-operational steering, composed of two independent steering engines, and fail-operational braking with two independent brake pressure sources.

Figure 2.5: Connectivity between vehicles [4].

- **Vehicle back-end.** This unit is in charge of data provisioning and collection for operator commands. It incorporates drive requests for initiation and supervision of vehicle movement. Additionally, it is in charge of data distribution for the provision of maps with static and dynamic information. Moreover, this unit monitors surrounding vehicles for documentation and map improvement. Finally, this unit incorporates a teleoperation system for rescue operations.

2.3.2. Communication systems

Communication systems are a fundamental element of AV operation. They are in charge of gathering necessary data and being able to process immediately, as well as developing a different way to transmit that information through both internal and external devices. They are the main objective of many current studies and research programs [5]. The following requirements set the scene to appropriate software development.

Intra-vehicle connectivity

There is an increasing number of sensors deployed in passenger cars. It is expected that some brands reach up to 200 per vehicle next year [5]. These sensors are required to report to the Electrical Control Unit (ECU) in real-time. The interest in intra-vehicle communication network design is consequently increasing. Wired solutions, such as a TTEthernet or a Controller Area Network (CAN), need cable connections between sensing elements and the ECU. These kind of networks imply a considerable weight to the vehicle [6], as well as add complexity to installation and maintenance. This fact, and the recent research and development of wireless sensor communication networks have allowed the application of wireless technologies in connected vehicles. There are several alternatives that have been widely investigated as Bluetooth, ZigBee, Passive RFID, Ultra-

Wireless solution	Bluetooth [23]	ZigBee [24]	Passive RFID [25]	UWB [26]	60 GHz mmWave [27]
Frequency band	2.4 GHz	868 MHz, 915 MHz, 2.4 GHz	915 MHz	3.1–10.6 GHz	57–64 GHz
Data rate	1, 2, 3 Mb/s	20–250 kb/s	<4 Mb/s	53.3–480 Mb/s	>1 Gb/s
TX power	1, 2.5, 100 mw	<1 mw	0	1mw/Mb/s	10 mw
MAC protocol	TDMA	CSMA/CA	EPCglobal	CSMA/CA &TDMA	CSMA/CA &TDMA
Modulation	GFSK (1 Mb/s) $\pi/4$-DQPSK (2 Mb/s) 8DPSK (3 Mb/s)	BPSK (868 MHz) BPSK (915 MHz) O-QPSK (2.4 GHz)	BPSK	MB-OFDM	Single carrier, OFDM
Application	Multimedia	Monitoring/Control	Monitoring/Control	Multimedia	Multimedia

Figure 2.6: Existing alternatives to wireless technologies [6].

Wideband or 60 GHz Millimetre Wave systems. Each of these technologies have different characteristics [7], as shown in Figure 2.6.

Inter-vehicle connectivity

The aim of this subsection is to briefly introduce the different inter-vehicle communication technologies. A schematic illustration of how each of these different mechanisms work is represented in Figure 2.7:

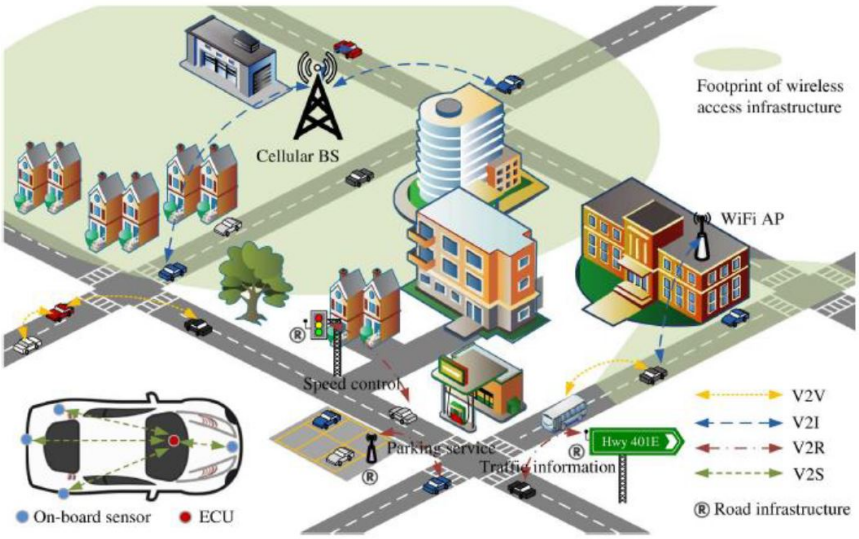


Figure 2.7: Inter-vehicle connectivity [6].

V2V → where vehicles exchange relevant information directly.

V2I[1] → where vehicles connect to the internet

V2R[2] → where vehicles exchange real-time information with the road infrastructure.

V2S → for intra-vehicle communication systems.

V2X → for vehicle to everything.

[1] Some references use the acronym V2N, Vehicle to Network.
[2] Some references use the acronym V2I, Vehicle to infrastructure.

V2V systems enable interaction with vehicles at close distance, within 300 meters range approximately. V2I enables internet access, but is mainly focused on access to multimedia platforms. V2R uses external devices located along the road infrastructure. The reception of real-time road information can play a significant role in the mitigation and avoidance of road crashes and their resulting effects.

2.4. Relevance of the AV in the project

It is only a matter of time before we start coming across more and more automated vehicles. Some vehicles have already reached level four of automation but there are still some issues that need to be solved in order to reach the ultimate level of automation. Technology and legal aspects are the main issues to be solved.

There are no convincing solutions yet for the anticipation of road user behavior, bad weather conditions and hidden objects. Moreover, there is no agreement on regulations, laws, rules for validations, ethics and liability. Finally, there are also unresolved issues like customer acceptance and trust [1] this is where this project comes into play. By knowing the reactions of drivers, human machine interfaces (HMI) can be developed in order to adapt to human acceptance. This project collects data from different driver profiles in order to construct a database to delve deeper into the driving behavior. The profiles are defined under the theory of the Big Five personality traits, which classifies the personality of an individual in five groups. The group that predominates from the others categorizes every person. There is a correspondence between the personality and the driving behavior. Therefore, with different profiles of drivers, their corresponding different driving behaviors can be obtained.

3

Personality theories and driving behavior

3.1. Personality theories

Understanding personality has always been a difficult and challenging task for psychologists; there are many theories that tries to provide the answers with different perspectives. Personality can be understood as the pattern of thoughts, feelings, actions, traits and behaviours that will lead to develop the character, temperament and nature [10].

The origin is also an uncertainty; some people believe that comes with the biology and genetics of a person, which remains constant through the years. Others believe that personality changes through life due to culture, experiences and environment. After many years of investigation, psychologists have developed many theories; some of them have become accepted and influencing. There are many definitions, views and approaches in order to classify the different personalities for the human beings [11]. The theories of personality, which helps to understand the patterns and behaviours, are fundamental when it comes to design new technologies. New devices are made to improve human life, for that reason, it's crucial that they make the expected impact and doesn't create negative or rebound effects.

The most influencing personality theories are developed through different perspectives and when placed in a business context, some are more useful than others. The most important theories that are accepted and currently influencing on the psychology field, define personality by psycho-dynamic or by traits. In psychoanalysis, Freud's theory is one of the most important and the one that has caused the biggest impact among

the psychologists. Nonetheless, it has been very criticized. Regarding definitions with traits the most important are the Big Five, Eysenck's theory and Cattell's 16PF [11].

In this project, a method to "quantify" personalities is required. The selection is based on the most representative volunteers of each pattern or behaviour so the results are the most accurate possible. After considering many approaches, the Big Five is the optimal for our purpose as discussed at TU Delft. The results are quantitative and the classification of the different personalities can be made quickly, this allows an objective and efficient selection.

3.2. Big Five personality traits

The Big Five personality traits, is a classification method based on the composition of five personality dimensions. These factors were found experimentally in an investigation of personality descriptions between different people. The big five is a descriptive model of personality, and psychologists have developed various methodologies to evaluate those five traits in an individual. The five main traits or factors are traditionally referred to as: factor O (Openness or openness to new experiences), factor C (Conscientiousness or responsibility), factor E (Extraversion), factor A (Agreeableness or kindness) and factor N (Neuroticism or emotional instability) [12].

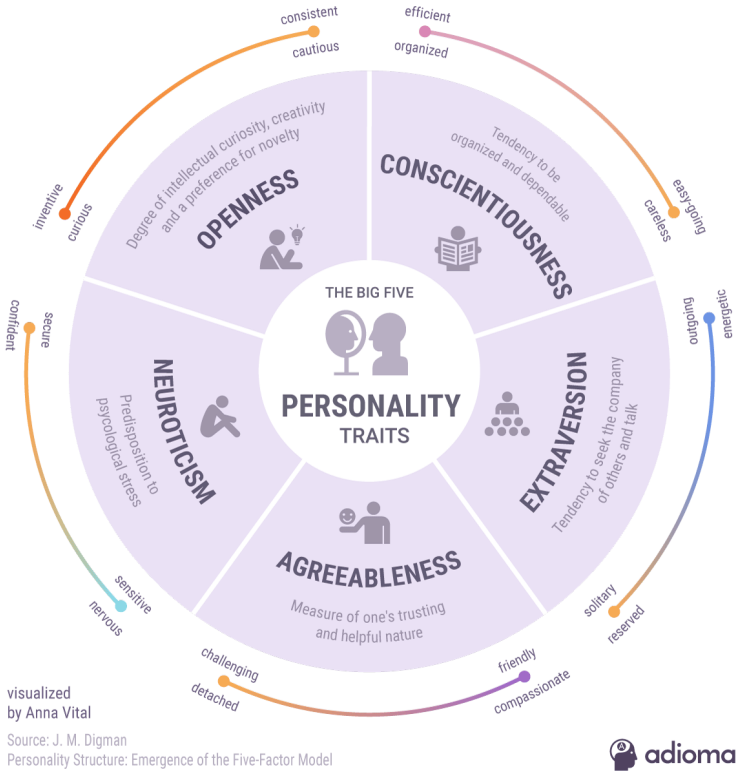


Figure 3.1: The Big Five groups and its characteristics [14].

3.2.1. History

Early research of the human traits date back to the 1930s by the scientific sir Francis Galton, he come up with the idea that the most socially relevant personality differences are coded in language through an adjective that designates them [13].

Subsequently, Gordon Allport and H. S. Odbert searched two of the most important English-language dictionaries words describing personality (states, traits, attitudes...). They finally reduced the list and obtained 4500 adjectives describing observable and relatively permanent traits. Afterwards, Raymond Catell reduced that list to 171 words and identified 16 personality traits after some tests made by him and his collaborators [13].

In the 1960s, Tupes and Christal used Cattell's personality traits, and found five recurrent factors. Norman replicated the work and discarded terms referring to states, and retained the traits. He also found that five major factors were enough to explain personality types. Norman called these factors Extraversion, Pleasantness, Consciousness, Emotional Stability and Culture [13]. Later in the 1990s, Goldberg gives the name of the Big Five and, finally, McCrae and Costa elaborate a questionnaire that consolidates and expands the model [13].

3.2.2. Descriptions of the particular personality traits

McCrae and Costa agreed that personality traits have two poles and represent a normal distribution. For each feature, most of the people have an intermediate score, and only a few people have extreme values [13].

Openness to experience (vs Conservatism)

The openness to experience factor distinguish people seeking new experiences from people who are better off in a comfort zone. People who systematically seek different and varied experiences will have high values in this factor. People who are open to experience tend to question traditional values, while people who are not open to experience tend to support traditional values and maintain a fixed lifestyle. People with high levels of openness are often creative, imaginative, curious and liberal and prefer variety. On the other hand, people who are not very open to experience are usually conventional, practical, conservative and not curious at all [12].

Conscientiousness (vs Negligence)

The conscientiousness factor describes the organized, self-contained, meticulous, ambitious, and self-disciplined people. In general, people with high values in this factor are hardworking, applied, punctual, persevering and tend to have a clear orientation of their goals. On the other hand, people with low values in the factor tend to be disorganized, negligent, lazy and tend to surrender to difficulties [12].

Extraversion (vs Introversion)

People with high values in the extraversion factor tend to be caring, jovial, talkative, sociable and fun. Extraverts like to be with people, and are often perceived as full of energy. On the other hand, people with low values in the factor tend to be reserved, silent, lonely, passive, and unable to express strong emotions. It doesn't mean that they are unfriendly or antisocial, rather, introverts are reserved in social events [12].

Agreeableness (vs Antagonism)

The agreeableness factor distinguishes compassionate people from cruel ones. People with a high level of kindness tend to be trusting, generous, complacent, condescending and friendly. People who approach the opposite extreme are often mistrustful, stingy, hostile, irritable, and critical towards others [12].

Neuroticism (vs Emotional stability)

People with high neuroticism values tend to be restless, temperamental, self-pity, anxious, impulsive and sensitive to stress. They tend to have problems related to their low self-esteem. People with low neuroticism values are usually quiet, gentle, impassive and are satisfied with themselves [12].

3.3. Other theories

3.3.1. Freud's Theories

Sigmund Freud (1856-1939) was an Austrian neurologist father of psychoanalysis and one of the greatest intellectual figures of the twentieth century. He started in the field of neurology and progressively divert to psychoanalysis, to which he devoted the rest of his life [15].

Freud's theories about personality attributes the mind to be the main responsible for both conscious and unconscious decisions, based on the psychic impulses. He believed that the basins for person's personality consisted in the interaction between the three aspects of the mind: the id, the ego, and the super ego. The three parts compete with each other for the decision-making [15]. According to Freud believes there is an interdependence between the three levels, which have their purpose on the development of the personality of an individual. The levels have the ability to resolve internal conflicts at specific stages of their development, which determine the future and maturity.

Id

The id is the primitive and instinctive component of personality. It is composed by the inherited components of personality including the instinct and the aggressiveness. It works impulsively and responds to our basic urges, needs and desires regarding of the consequences [15] [16].

Ego

In accordance with Freud, the ego is the decision-making component of personality. Ideally, the ego works by reason, whereas the id is chaotic and unreasonable. Both id and ego seek pleasure and avoidance of

pain, but the ego if focused on realistic strategies to obtain the pleasure. The ego works rational, realistic and orientated towards problem solving. It helps to control their impulses and demonstrate self-control [15] [16].

Super Ego

The super ego functions at a conscious level and incorporates the values and morals of the society, learned from parents and environment. Its function is to control the Id impulses, which are forbidden by the society. It persuades the ego with moralistic values rather than realistic ones. The superego consists of two systems: The conscience and the ideal self; causing feelings of guilt or proud [15] [16].

Figure 3.2: Representation of the three aspects of the mind [15].

3.3.2. Eysenck's Theory

Hans Eysenck (1916-1997) was an English psychologist of German origin, specialised in the study of personality. Eysenck's theory is based primarily on physiology and genetics. He believed that the personality traits of the human behaviour were grown out of our genetic inheritance. The ability to learn and adapt to the environment is affected by the type of nervous system that is inherit. He was interested on the temperament, which is an aspect of the personality that is genetically based. However, he did not exclude the possibility that some aspects of personality are learned [17].

He found that behaviour could be represented by two dimensions: Introversion/ Extroversion and Neuroticism/ Stability. Each aspect of personality is related to a different biological cause. Personality is dependent on the balance between excitation and inhibition. This is a process of the autonomic nervous system, which is a control system that acts largely unconsciously [18].

Extraversion/introversion

Extraverts are sociable and crave excitement, carefree and optimistic. They prefer more stimulating environments. Eysenck believed that this was because they inherit an under aroused nervous system, so they seek for sensations in order to restore the level of optimum stimulation. On the other hand, introverts are on the other side of the scale. They are quiet, reserved, plan their actions and control their emotions; they tend to be serious and pessimistic. As opposed to the extraverts, they have an over-aroused nervous system so they tend to avoid sensations and stimulation [18].

Neuroticism/stability

Someone who has high levels of neuroticism, is more unstable, tend to overreact to the situations and may be quick to worry, to get angry or to be frightened. They are overly emotional and find it hard to calm down once upset. Their nervous system has a quick response towards stress. A stable person is generally less reactive to stressful situations and can remain calmed. They have a low reactivity of their sympathetic nervous system [18].

According to Eysenck, both dimensions of neuroticism/stability and extroversion/introversion form a big variety of personality characteristics.

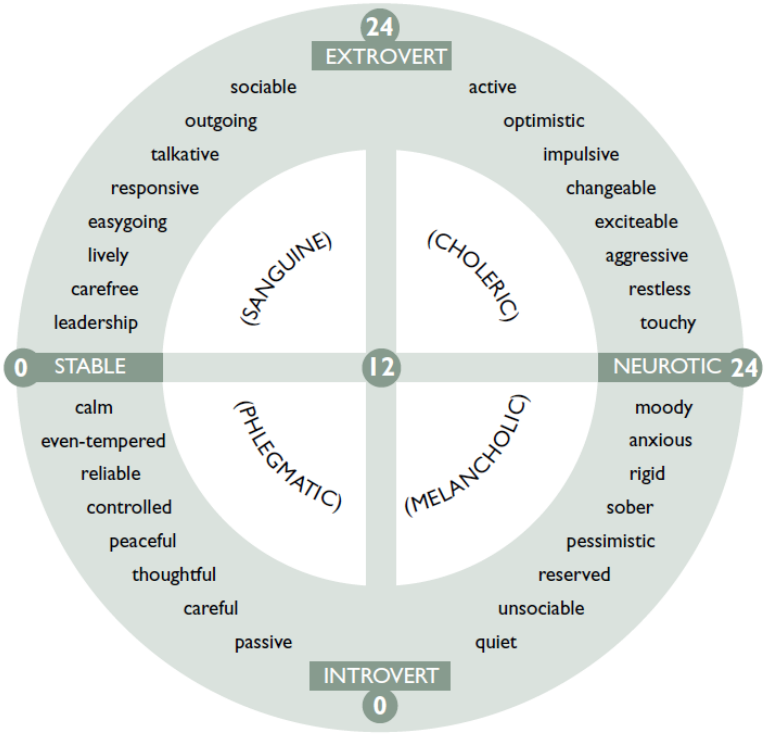


Figure 3.3: Different characteristics of both dimensions [19].

3.3.3. Catell's 16 Personality Factors

Raymond Bernard Cattell (1905-1998) was a British psychologist from the XX century. Cattell theorized about intelligence and made his own model about personality. Unlike Eysenck's bi-dimensional theory, Cattell describes how personality works with 16 Personality Traits. He believed that the differences between the personalities of individuals are described and explained by 16 different categories. He thought that personality was something that could be studied and organized. He elaborated a description of the underlying personality traits [20].

Cattell collected data from a different range of people through three different sources and identified 16 personality traits / factors common to all people. He produced a personality test that measured each of the sixteen traits. The 16PF (16 Personality Factors Test) has 160 questions in total, 10 questions relating to each personality factor [20].

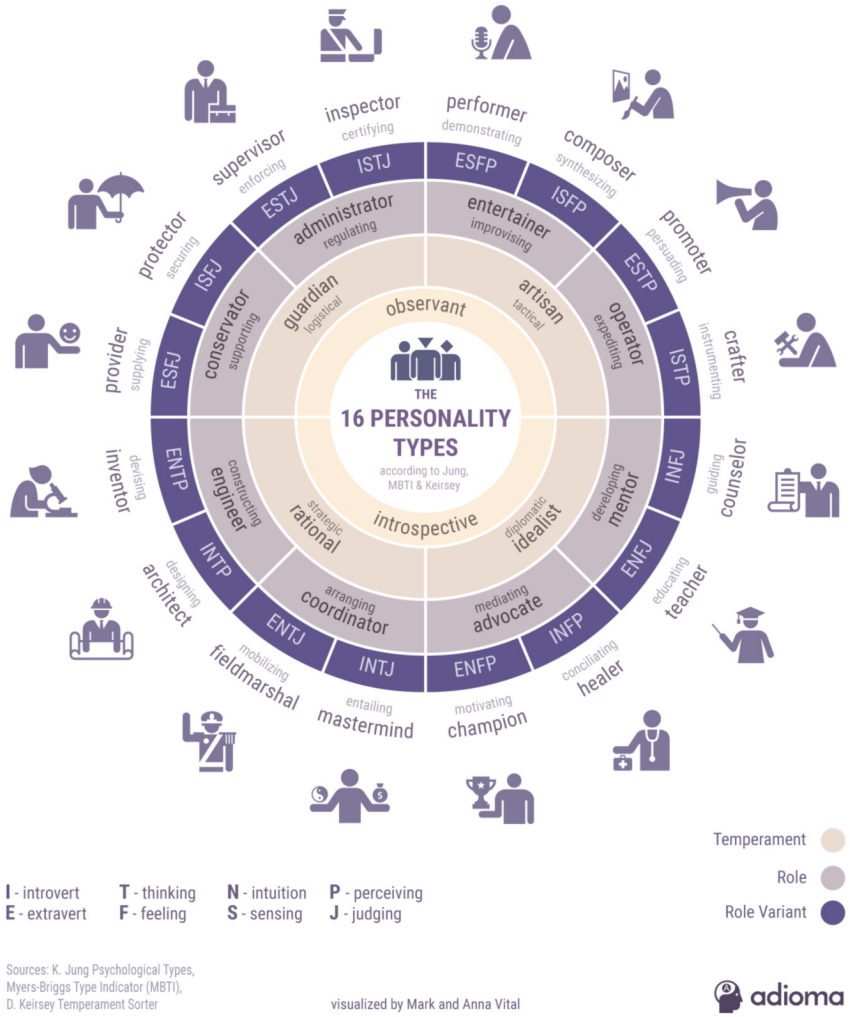


Figure 3.4: The 16 Personality types [21].

3.4. Correlation between personality and driving

As it is said in the previous chapter, there are many ways to define and classify someone's personality. This can also be extrapolated into driving. This chapter comes closer to the behaviour in traffic and the correlation between the personality and driving.

3.4.1. Behaviour in traffic

Although many people are able to drive a car safely after some training, human failure and user behaviour plays a major role in significant percentage of crashes that are produced in traffic. Because of that, its study is so important for improving safety on roads. Hazard perception is one of the most important aspects when driving that can differentiate between novice and experienced drivers [8]. That is to say, experience is a key element to take into account regarding response to danger on the wheel.

Perception of danger has an important role on behaviour, for example, road crashes are less feared than fatalities from plane crashes, even though there are many more deaths on the road per year. Road users tend to fear spectacular and unlikely events. Some of the following facts could help to make an approach of why people behave in that way [22]:

- **Fear strengthens memory**. Events that are infrequent but frightening as a terrorist accident tend to be overestimated. Otherwise happens with ordinary events, that have an underestimation even though are also dangerous.
- **It is a common thing**. As we see accidents almost every day in the media, we tend to see them as a common thing and no as a dramatic event.
- **Illusion of control**. People gain confidence when they believe to be in control of the situation. For that reason, many people fear to take a plane but not to drive a car even there are many crashes per day.

Apart from that, as the number of vehicles is increasing on the road, more demand of processing information is required. Reaction to stimulus depend on the mental workload, which is the amount of information capacity that is used for task performance [23]. Performance and workload are related to task demands and can be correlated as the shown in Figure 3.5.

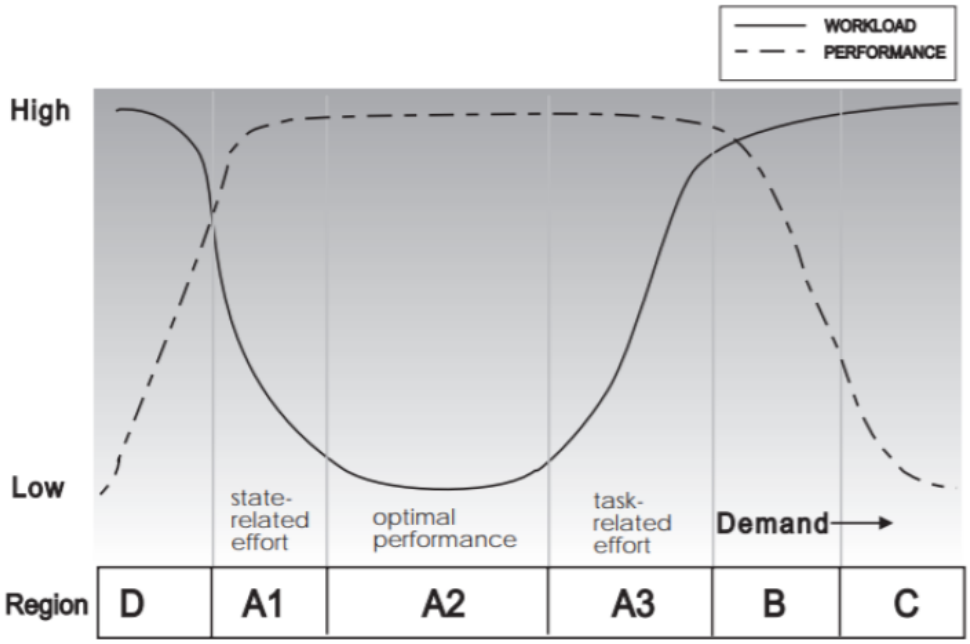


Figure 3.5: Model of mental workload, task performance and demand [23].

On Region D, the demand starts and so does the user's status. A2 is the region where the performance is optimal: the user can easily cope with all the requirements and reach a high level of performance. In regions A1 and A3 performance keeps high but the user has to make an effort to accomplish all the task requirements. In region B, demand starts to be so high that the workload increases which make the performance decline. Finally, in region C, the user is overloaded with the demand and that makes a minimum level of performance.

In order to successfully control or avoid hazards, the capability to complete the production task is needed. The difficulty of the task is indicated by the task demand, which include both physical and cognitive demands. As the task demand increases, more is the chance of error and loss of control. It can be increased due the number of elements that have to be controlled and the difficulty to control these elements.

The variables of task demand and applied capability are shown below in the Figure 3.6 [24]:

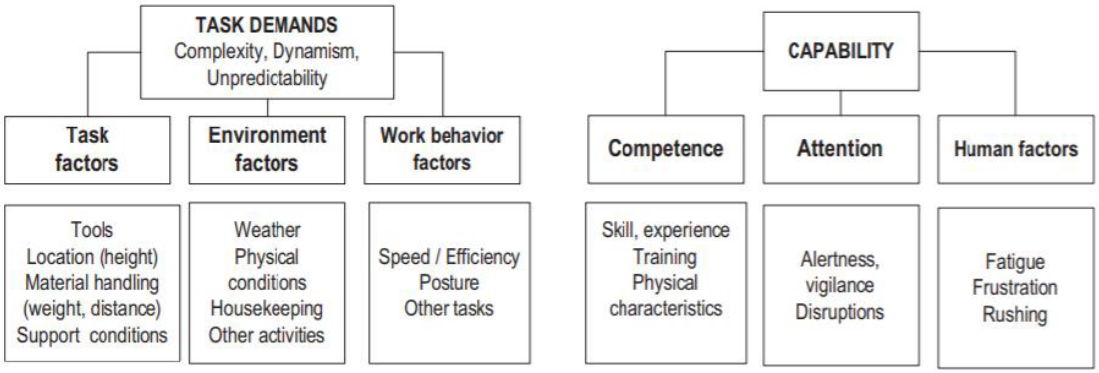


Figure 3.6: Variables of task demands and applied capability [24] .

This model proposes three categories for the task demands: task factors, environmental factors and work behaviours. These factors will determine the difficulty of the performance and the possibility to make errors or have a crash.

On the other hand, applied capability determines the ability to cope with the task demands and it depends on the competence of the user, the level of attention and the human factor. The competence of the user is increased with the experience, it can be trained and includes skills and physical condition, familiarity with the situations, reaction time etc. The attention given to the task and the hazard is a limited resource, multitask demands reduce the attention to any single demand which can lead to a loss of attention from the hazard. Attention to the production task, can also act as a distraction from the hazards. Human factors can also reduce competency, the four states of mind that can lead to this situation are rushing, fatigue, frustration and complacency [9].

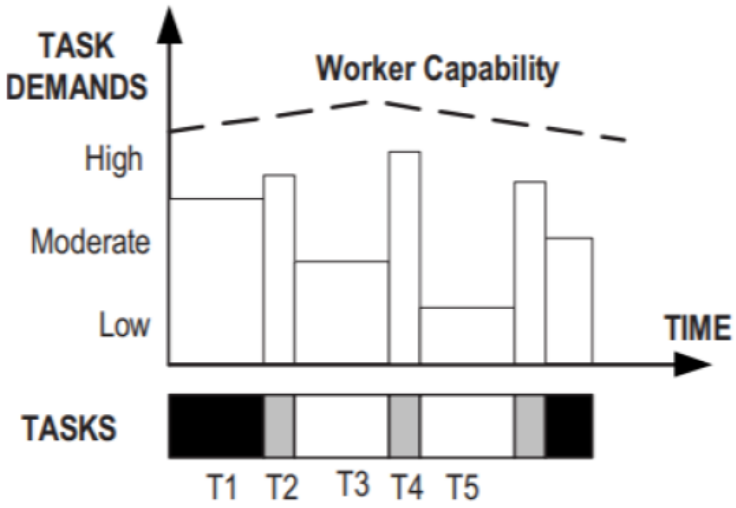


Figure 3.7: Work situations described by the tasks, task demands and capabilities [9].

3.4.2. Personality and driving

The correlation between personality traits and driving behaviour as well as accident involvement is an issue that has been studied for many years. One of the goals of the autonomous vehicles is to avoid the errors made my humans. The most common errors can be quantified in a table, but its not as clear which profile of driver is more likely to be involved in car accident. There are many studies made with different methods which prove clear evidences of the correlations.

Philipp Yorck Herzberg (2009) applied a prototype approach based on the Big Five personality traits model in order to obtain empirical evidence for a consistent relation between personality and driving behavior. The data of two samples of drivers was analysed and clusters of the results revealed three prototypes in both groups. The prototypes, labeled Resilient, Overcontrolled, and Undercontrolled, were found to be different with the reliably in accident involvement and driving behavior, indicating differential accident liability. The results of the samples indicated that the group of undercontrolled Undercontrollers were the most problematic drivers followed by Resilients, whereas Overcontrollers most consistently obeyed traffic rules and drove accordingly [25].

A research made by Bahram Esmaeili (et. all, 2012) investigated the relationship between personality traits of drivers and the number and amount of fines they have in a year. The study was carried out on 120 male taxi drivers in the south of Iran. Predictive variables were the Bif Five personality traits, the criterion variables were the number and amount of fines the drivers have had the last three years. The result of regression analysis showed that the factors were able to predict (negatively or positively) the number and amount of financial fines the drivers had during the last three years [26].

Years later, Biying Shen,Weina Qu et all. (2017) made a research about the relationship between personalities and self-report positive driving behavior in a Chinese sample. The main goals of the study were to explore the positive driving behaviour and its relationship with personality in a Chinese sample. A total of 421 licensed drivers completed many questionnaires and the results showed that the Positive Driver Behavior Scale (PDBS) was siginificantly correlated with the Big Five Inventory. The PDBS was positively correlated with extraversion, agreeableness, conscientiousness and openness and negatively correlated with neuroticism [27].

There are also many previous studies which used the Big Five Inventory to predict many types of driving behaviour. Arthur and Graziano (1996) demonstrated a significant inverse relation between Conscientiousness and at-fault accidents or moving-violation tickets [28]. Other study, a negative correlation between car accidents and Conscientiousness was found (Arthur & Doverspike, 2001) [29]. Moreover, Ulleberg and Rundmo (2003) showed that traits of Extraversion, Agreeableness and Neuroticism were significantly related to risk-taking behavior [30].

The findings and correlations can be useful in recognizing high-risk drivers. Such information could be important in companies or offices related to driving. Also to inform the drivers about their personality and it's relation with their accident rate so they are able to take action if needed.

To sum up, there are many articles that, by means of empirical evidence, demonstrate a correlation between the driving behaviour and the personality. This project, as part of TU Delft research, uses the Big Five Inventory to find representing candidates of each category in order to analyse their driving behaviour with autonomous vehicles in a simulator. Not only is an easy way to quantify the personality, but also fast. The main goal is to gain knowledge with the coexistence with autonomous vehicles, to be a step closer to make this new and auspicious way of driving a reality.

4

Selection based on the Big Five Scoring

TU Delft is undertaking a project to investigate the behaviour of drivers of automated vehicles based on the Big Five Personalities. As mentioned in the previous chapter, many investigations relate the personality with the driving behaviour. The goal of the TU Delft project is to study the driving patterns in relation to the personality. In order to conduct this project, representative participants must be recruited. The experiment will be performed in a driving simulator located at TU Delft.

4.1. Participant selection

The selection of participants for the driving simulator experiment is made through a variety of stakeholder employees and/or participant pools. The data is analysed and then the most suitable participants to partake in the experiment are selected. The criterion that is predetermined to select the participants is their score in the Big Five Inventory for each category.

4.1.1. Making the questionnaire

The questionnaire, attached in the appendix A and B, collects different types of responses: a consent form, availability and demographic questionnaire and finally the big five Inventory. In order to reach a wider audience the questionnaire is in English and in Dutch.

The consent form gives information about the privacy of the responses and is essential when it comes to the collection of sensitive data. Any information provided is handled with the utmost care, and privacy

is given priority at all times. None other than the researcher handles the information, which in question, is henceforth considered to be responsible, and thereby acknowledges his or her responsibility. The information is only used for the purpose of the study. All directly identifiable data, such as names and email addresses, that are collected will be deleted after the experiment has been completed. Furthermore, the availability and demographic questionnaire collects data such as the age, gender, education and driving information. It also asks for the drug use, which is considered a sensible data.

Finally, the Big Five Inventory, which is composed for 44 questions about someone's personality traits. It has to be answered with a number from 1 to 5, depending on how you identify with the statement:

- 1 → Disagree Strongly.
- 2 → Disagree a little.
- 3 → Neither agree nor disagree.
- 4 → Agree a little.
- 5 → Agree Strongly.

4.1.2. Selecting the candidates

The selection of the participants is made through a selection algorithm. The answers to the questionnaire are first introduced in an excel file. Then, this file is read by a Matlab code that generates the lists of the selected participants who have the highest results in each category. It is possible that one participant acquires the highest score or the same percentage in more than one category. In order to make the optimal classification and make sure the right participants for each category are chosen two algorithms are used.

Selection algorithms

Two algorithms have been designed, one works better with reduced number of participant and the other does otherwise.

Selection algorithm 1 gives back the exact number of participants that are needed, with the optimal volunteers for each category.

Selection algorithm 2 gives the optimal volunteers for each category but without taking into account the number of participants needed. This code is more accurate because it gives you back the real category in which every person belongs, but it only works with a large number of participants, because in reality, the categories are not uniformly distributed.

For each category both selection algorithms compute a combination of the score of all the answers with some reverse-score negatively keyed items. The specifications are attached in the Annex C and D. Every individual obtains a score for every category and are sorted according to their respective results.

From this point forward, the selection algorithms are different. Selection algorithm 1 takes the X first participants for each category. If the same participant appears in the top X of more than one category, he or she is only retained in the category where the score is maximum. The gaps that appear are filled with the remaining following participants from each ordered list.

Selection algorithm 2 selects the highest score for each individual. This way, each individual appears in one list only. The selection algorithms are included in the Annex F and G.

4.2. Results

The total number of participants is 86, between the ages of 23 and 66, among which 34 are women and 52 are men.

Figure 4.1 shows the distribution of the scores of the candidates, together with the maximum and minimum possible scores for each category. The average of the scores is indicated with a red circle.

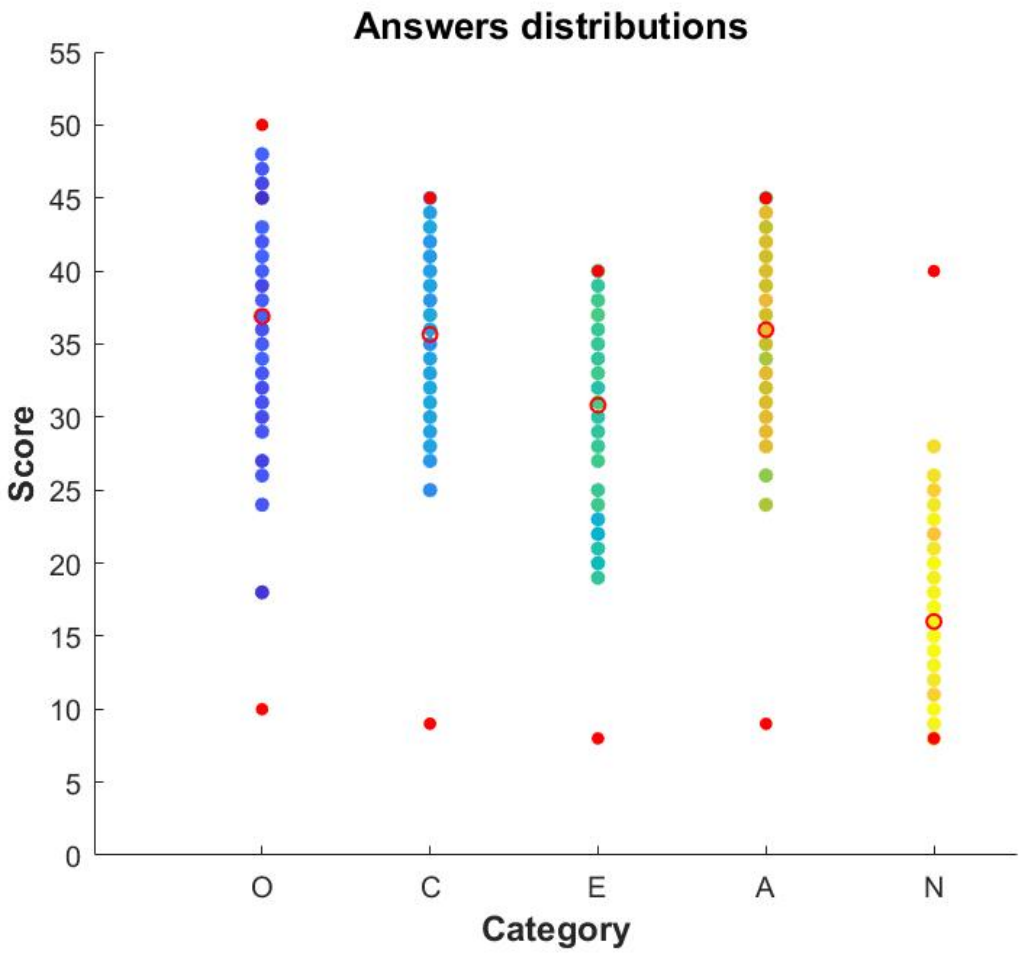


Figure 4.1: Answers distributions of the candidates. O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness and N:Neuroticism.

It can be observed that, in categories C, E and A the maxima are reached, in contrast to category N where no participant scores the maximum. In categories O, C, E and A no participant is close to the minimum whereas in category N the minimum is reached.

On the other hand, the category O has the scores more spread as compared the others, but does not reach the minimum and maximum. This means that is the category that contains more information regarding personality. In order to compare the different scores of the participants, we decided that the answers needed to be normalized.

Table 4.1 shows the mean and the standard deviation for each category:

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Mean	36.895	35.663	30.814	35.977	16
Standard Deviation	5.535	5.068	5.3281	5.068	4.4458

Table 4.1: Mean and Standard Deviation of categories.

To normalize the answers the mean needs to be substracted from the scores and the resulted value divided by the standard deviation.

The following figures show the different histograms for the categories, composed with the scores obtained by the participants.

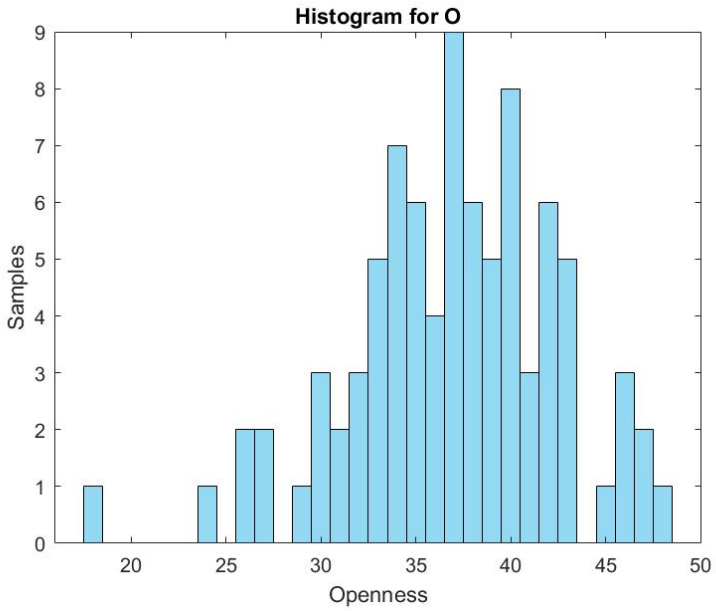


Figure 4.2: Histogram for category Openness.

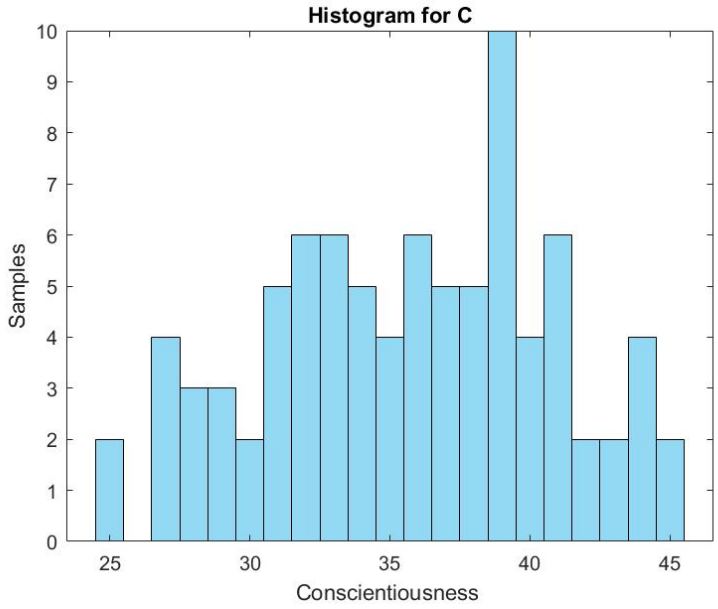


Figure 4.3: Histogram for category Conscientiousness.

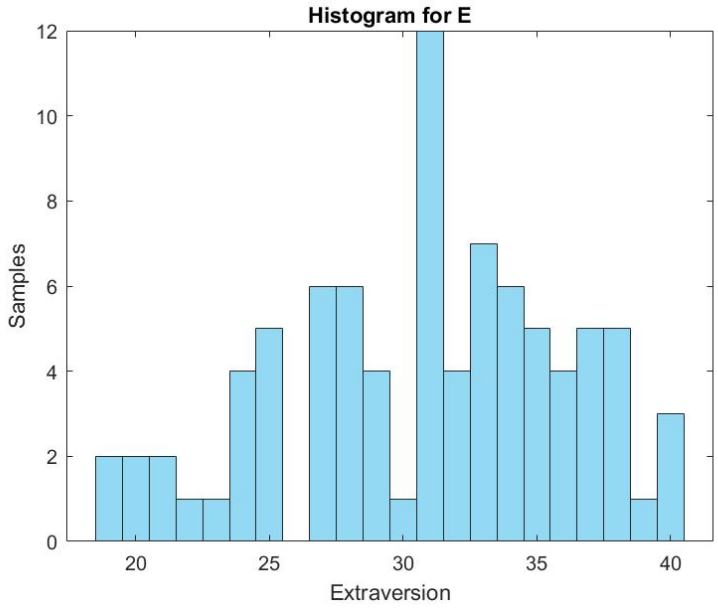


Figure 4.4: Histogram for category Extraversion.

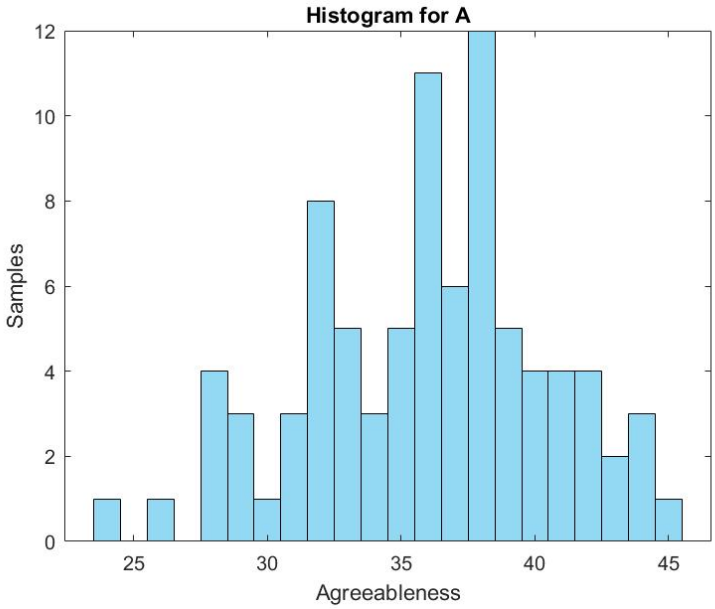


Figure 4.5: Histogram for category Agreeableness.

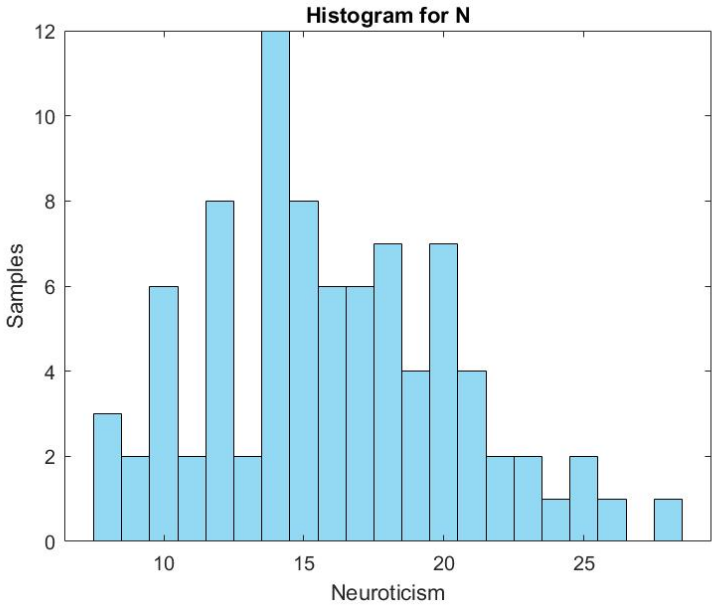


Figure 4.6: Histogram for category Neuroticism.

The histograms of the scores show that there is only one person in the maximum for categories O, A and N. For category C and E there are two and three individuals respectively. In the priority of the selection, these details are taken into account. If the same person is in the top list for the category E and A, it would be better to retain this person in list A, and in list E, choose another individual who are also on the top. This is taking into account in the selection algorithm 1.

For selection algorithm 2, which takes into account the maximum score of each person after normalizing

the answers, the number of candidates are the following:

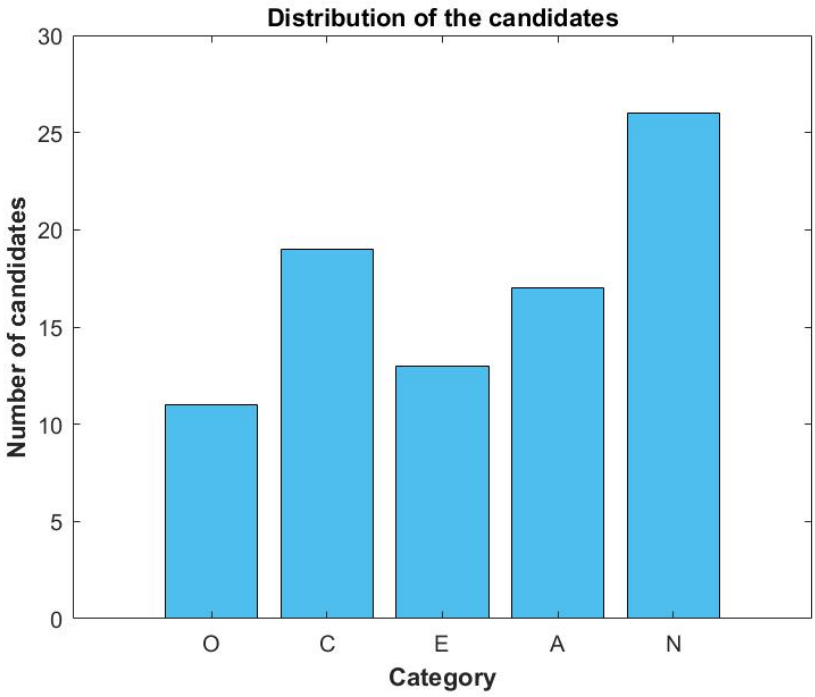


Figure 4.7: Distribution of the candidates with the Selection algorithm 2.

Selection algorithm 2 might not be the best option if the selection requires a minimum number of candidates since there is no guarantee that the minimum will be attained. Even though this algorithm might not give back an exact number of candidates for each category,it is a better method of selection. It returns candidates with the highest scores in the categories. Selection algorithm 1 however might be more practical in some cases since it always provides a list with the predetermined number of proper candidates.

The lists of the candidates using both lists, taking into account the top 10, are the following:

Selection algorithm 1				
Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Aranka V.	Krista W.	Lan van R.	Pascal V.	Colin K.
Sylvia V-M.	Eveline van O.	Eric van O.	Roger V.	Caroline S.
Frans K.	Marleen D.	Iris G.	Wim ten B.	Irene H.
Duan T.	Michael M.	Ghada A.	Yvonne van de S.	Jasper T.
Martin K.	Arnolda van V.	Marieke de B.	Marieke H.	Femke van de W.
Matthijs H.	Ron C.	Jaap B.	Costyn van D.	Regina van der W.
Adagonda V.	Dennis de K.	Tjitse van den B.	Henk E.	Marjanne G.
Maurits van B.	Erwin de L.	Patrick S.	Jolanda de F.	Mirjam van H.
Eskandar Z.	Elleke H.	Xandry A.	Herman W.	Andres de G.
Martijn R.	Hannie G.	Ed N.	Chris K.	Arlene T.

Selection algorithm 2				
Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Aranka V.	Krista W.	Lan van R.	Pascal V.	Colin K.
Sylvia V-M	Eveline van O.	Eric van O.	Roger V.	Caroline S.
Frans K.	Marleen D.	Iris G.	Wim ten B.	Irene H.
Duan T.	Michael M.	Ghada A.	Yvonne van de S.	Jasper T.
Matthijs H.	Arnolda van V.	Marieke de B.	Marieke H.	Femke van de W.
Adagonda V.	Ron C.	Jaap B.	Costyn van D.	Regina van der W.
Maurits van B.	Dennis de K.	Tjitse van den B.	Henk E.	Marjanne G.
Eskandar Z.	Ed N.	Martin K.	Jolanda de F.	Mirjam van H.
Martijn R.	Erwin de L.	Xandry A.	Patrick S.	Andres de G.
Johan H.	Elleke H.	Martin de B.	Herman W.	Arlene T.

Figure 4.8: Lists of the candidates using the two selection algorithms.

Both lists have similar results but some candidates are placed in other positions or in different categories. Nonetheless, the results are pretty similar and in the case of only needing 10 candidates both lists would work properly. If the number of candidates of each category needs to be higher than 11, Selection algorithm 1 might be the best option since it gives back an exact number and the results are proved to be similar as the selection algorithm 2.

5

Selection based on Principal Component Analysis

The main goal of this chapter is to obtain a new classification of the subjects under analysis based on the mathematical information contained in the questionnaire. Many methods can be employed for this task, this project uses the Principal Component Analysis in order to identify which linear combinations of answers are more significant. The main idea of PCA is therefore to reduce the number of the dimensions representing the data provided by the questionnaire. For example, in this case, there are 44 questions in our questionnaire and the answers of each subject are consequently represented in a vector of 44 dimensions. Careful analysis of the answers reveals however that not all the questions provide the same amount of information and many of them might be highly correlated. PCA provides a new set of coordinates which are mutually uncorrelated and where the amount of information that they contain can be quantified. Therefore, with only the principal components classification, most of the information can be retained.

5.1. Introduction to the PCA

Principal Component Analysis, or PCA, is one of the most important dimensionality reduction methods that is currently used. Dimensionality reduction is used when a big amount of data needs to be analyzed and only the most important features are kept. Principal component analysis is a procedure which converts a set of correlated variables into another set where variables become uncorrelated. It performs an eigenvalue and

eigenvector decomposition and the amount of information contained by each of the new variables can be quantified by the value of the corresponding eigenvalue. Figure 5.1 shows an example of a PCA plot where the data is represented using only the two principal components differentiating 3 types of Iris.

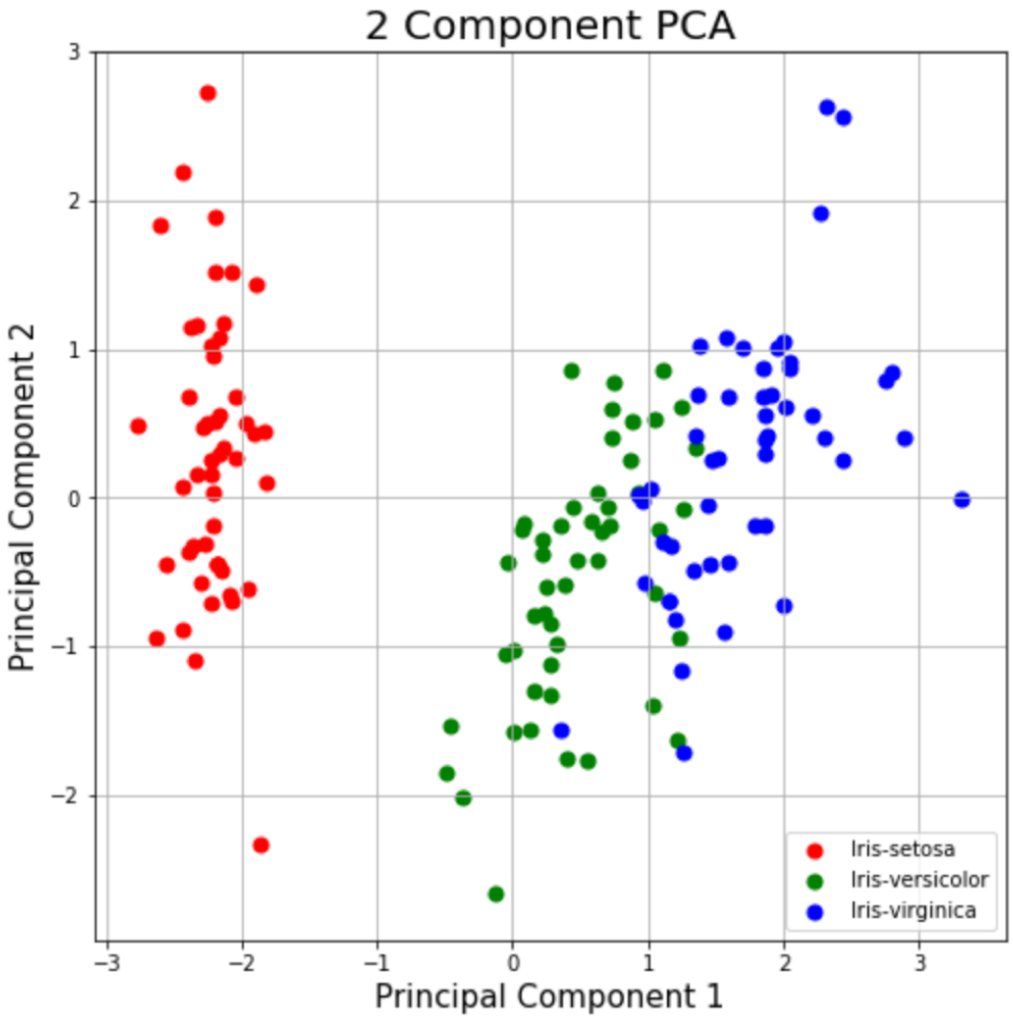


Figure 5.1: PCA example [31].

As explained before, there are many personalty theories based on psychological analysis, but in general mathematical analysis is not considered. This project uses PCA to study the information that the Big Five Inventory gives. It analyses the 44 questions based on the answers of the candidates, and makes a reduction taking into account their relevance. The main idea is to identify which are the linear combinations of questions that provide most of the information. These combinations are called principal components of the data.

5.2. PCA

5.2.1. General Insight

As said before, PCA is a method used in order to analyze a big amount of data. It reduces the number of variables and keeps the highest possible amount of information.

For example, with the information of the transcription of different gens in two cells, the correlation between them can be seen in a plot [32]:

	Cell 1	Cell 2
Gen 1	3	0.25
Gen 2	2.9	0.8
Gen 3	2.2	1
Gen 4	2	1.4
Gen 5	1.3	1.6
Gen 6	1.5	2
Gen 7	1.1	2.2
Gen 8	1	2.7
Gen 9	0.4	3

Table 5.1: Transcription of the gens in two cells.

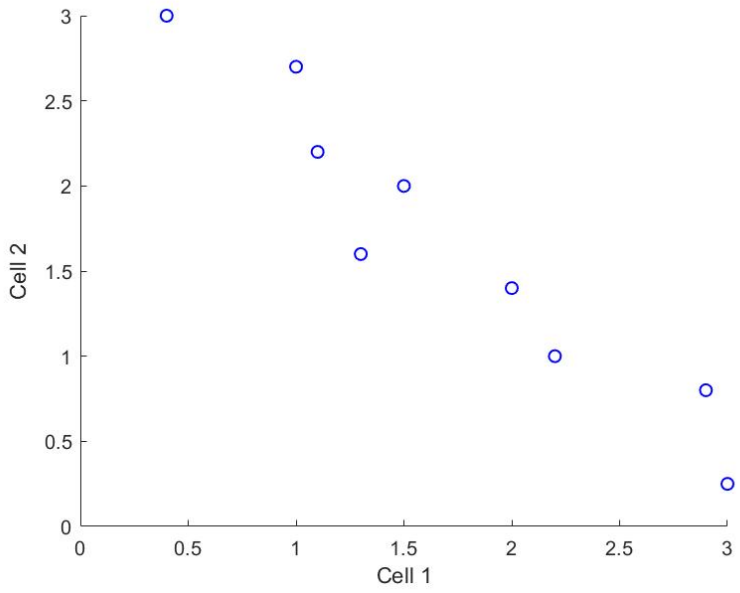


Figure 5.2: Example of a negative correlated cells.

The blue circles are the different gens and the axes are the transcriptions of the two cells. It can be observed that there is a negative correlation between the transcribed gens in each cell; it means that probably

they are a different type of cell since they have different transcriptions. In a 2-D plot, is easy to identify if there is a correlation between them, if there is a negative correlation like the example or they simply do not have any kind of correlation.

The comparison with all the cells can be done two by two, but it would take a long time and would be difficult to identify the correlations between many of them. An alternative much easier, faster and effective is drawing a Principal Component Analysis plot. A PCA plot converts the correlations, or the lack of them, among all of the cells into a 2-D graph like the following:

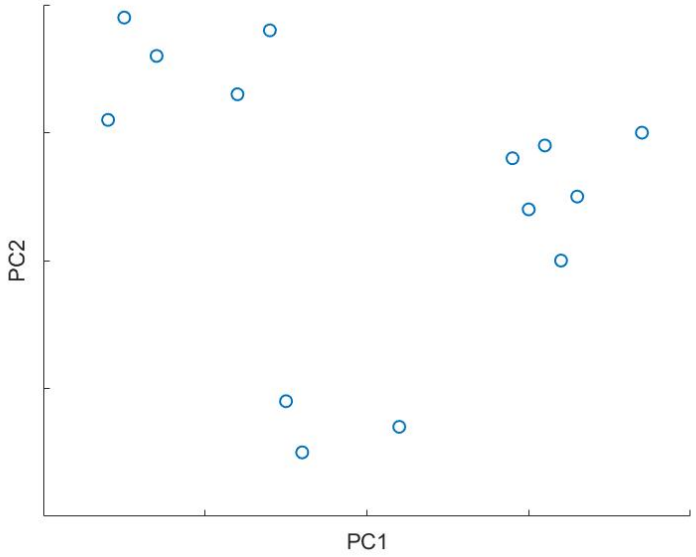


Figure 5.3: Example of a PCA plot.

Cells that are highly correlated cluster together, in the example there are three groups of cells based on the information of their genes. It also tells which gene (or variable) is the most valuable for clustering the data. The PCA plot not only distinguish the different clusters but also rank them in order of importance. Differences along the first principal component axis (PC1) are more important than differences along the second principal component. Principal components also tell how much information they contain, by the corresponding eigenvalue.

5.2.2. Mathematical description

After a general idea of the PCA and how it can be used, this subsection introduces the mathematical formulation. First, with an example of two variables and then in general [33].

Two variables example

With a dataset of only two variables, the first step is to subtract the mean value, so the transformed variables are zero mean. By doing so, the points get located around the origin of coordinates. The next step is to comput the covariance matrix:

$$\Sigma = \begin{bmatrix} \text{Var(X)} & \text{Cov(X,Y)} \\ \text{Cov(X,Y)} & \text{Var(Y)} \end{bmatrix}. \tag{5.1}$$

From this matrix, two eigenvectors and two eigenvalues are extracted. In Figure 5.4, the eigenvectors are represented by a green and a red line, the respective eigenvalues give their length. They not only give the direction in which the samples are spread, but also its variance along that direction.

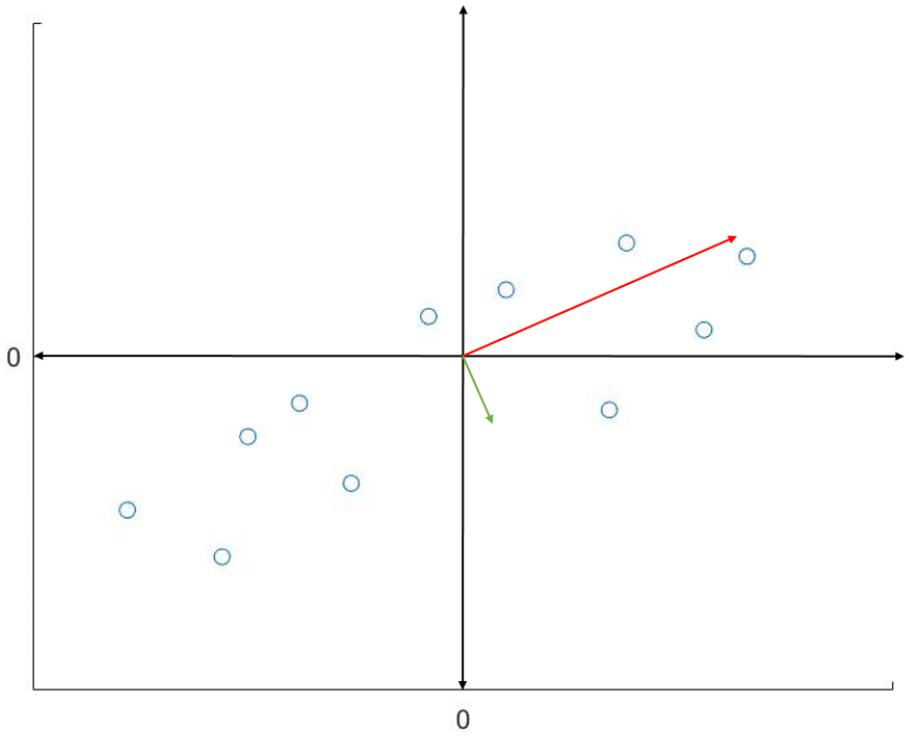


Figure 5.4: Eigenvectors of the sample.

The following step is to project all the samples in the direction that has the largest eigenvalue. By doing so, the number of components is being reduced to one, which is the most representative of the information contained in the samples.

General application

In the case that there are more than two variables, for example N, the procedure is the same. First of all, the covariance matrix must be created, which will be an NxN matrix. Then the eigenvectors and eigenvalues of the covariance matrix are calculated. The eigenvalues of the covariance matrix are real and non-negative since the matrix is symmetric. Eigenvalues are sorted from larger to smaller and the same order is applied to their corresponding eigenvectors. If the objective is only reduce dimensions to say M, keep the M biggest. Otherwise, if the objective is to represent the samples in two or three dimensions, keep the two or the three biggest respectively.

For creating the plot, the next step is to project the dataset using the selected eigenvectors. For instance, a 2-D plot, could be a good representation of the N variables if the two biggest eigenvalues are much bigger than the rest.

In fact, the information that gives each component is given by the formula:

$$I_i = \frac{\lambda_i}{\sum_{j=1}^{N} \lambda_j}, \tag{5.2}$$

where λ_i is the eigenvalue of the i^{th} component and N is the total number of eigenvalues. By adding the information given by each component it can be seen how much information the PCA of M components provides, compared to the original:

$$I_{PCA_M} = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{j=1}^{N} \lambda_j}. \tag{5.3}$$

To illustrate the use of this parameter a couple of examples are shown in the Figures 5.5 and 5.6. It can be seen how much information each component has. In Figure 5.5 more than the 90% of the information is contained in the first two components. However, in Figure 5.6, PC3 and PC4 account for a substantial amount of information.

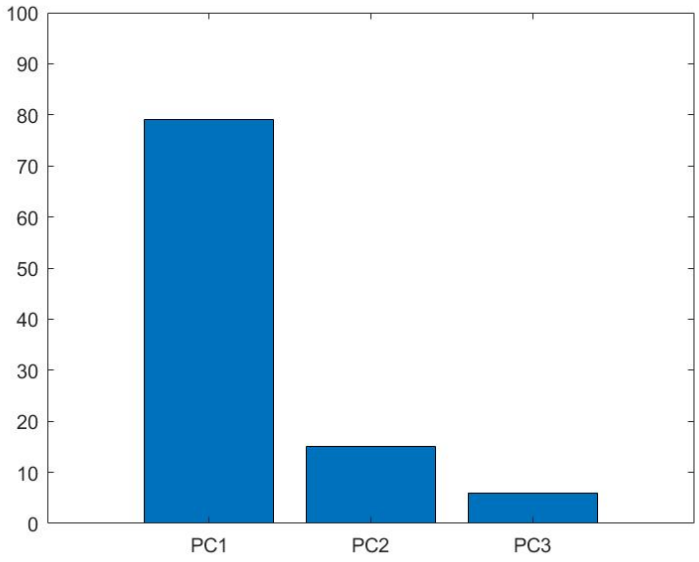


Figure 5.5: Principal Components that give good information.

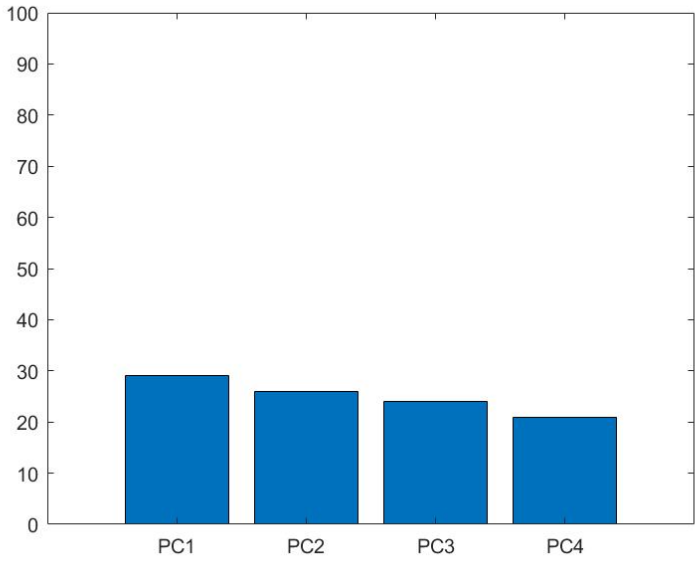


Figure 5.6: Principal components with less accuracy.

5.3. PCA Applied to the Big Five Inventory

5.3.1. Results of the PCA

This part uses the PCA methodology in order to mathematically analyze the dataset of the answers to the Big Five Inventory.

In this case, in addition to the 86 people who made the questionnaire for TU Delft, 34 answers have been

added from volunteers from Barcelona, thus increasing the number of subjects to 120.

First of all, the variance of the answers of each 44 questions is illustrated in Figure 5.7:

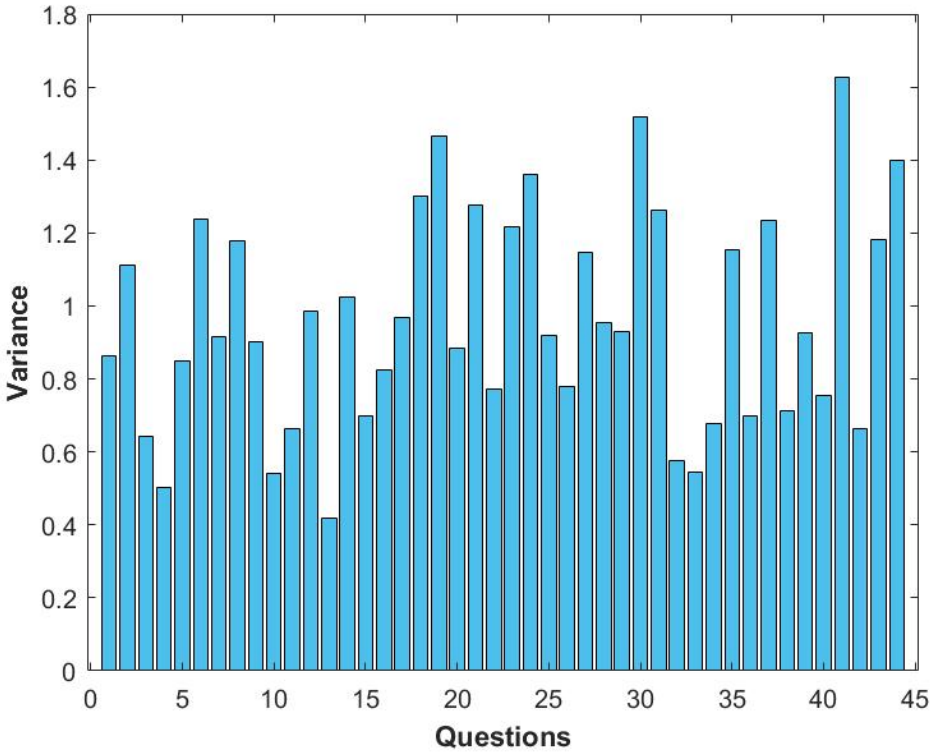


Figure 5.7: Variance of the 44 questions.

This allows to obtain a general overview of the information that the answers contain. The bigger the variance, the highest the amount of information as well as the relevance of the question. For instance, if every individual gives the same answer to a question, this question is useless (ans has a zero variance).

Once the PCA is applied, the distribution of the two first principal components of the 120 individuals is shown in Figure 5.8:

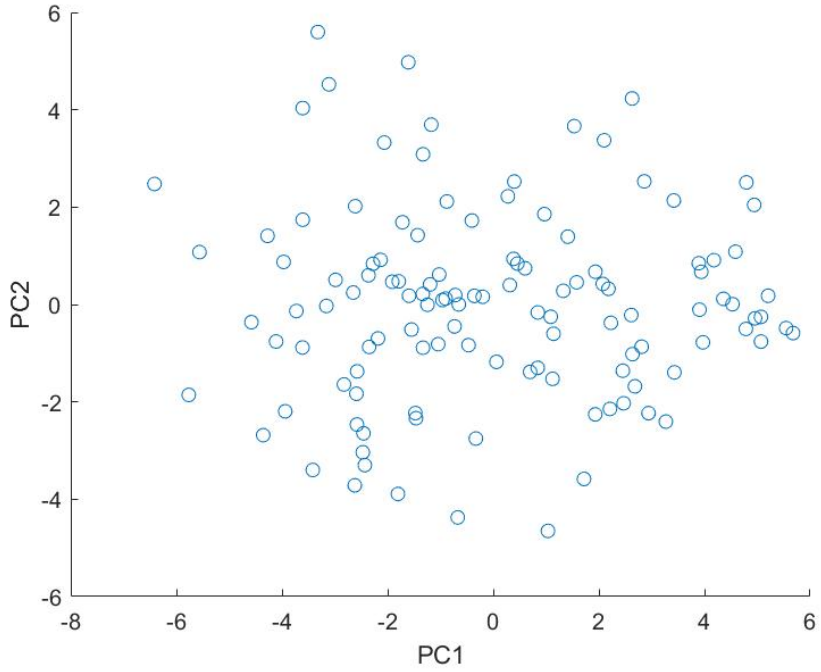


Figure 5.8: Plot of the PC1 and PC2.

Figure 5.8 shows a distribution of the 120 samples through both axes. In general, the features of the candidates are scattered and not accumulated at the extremes. It can be seen that there are no clusters. It means that the personalities of the candidates are all different from each other without following any pattern.

Nonetheless, this doesn't mean that the PCA is wrong, it only says that the distribution of the personalities among the candidates is homogeneous.

The following plot shows the amount of information contained in every component.

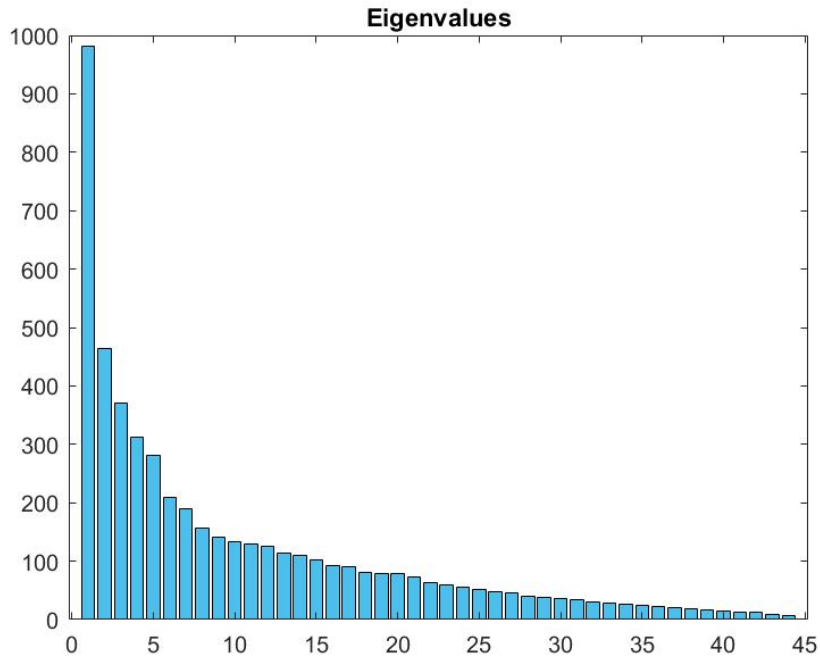


Figure 5.9: Distribution of the eigenvalues.

It can be seen in Figure 5.9 that the first eigenvalue contains a considerable amount of information as compared to the others.

The projection on the first PC1 contain the highest amount of information and it is also considerable higher than the second.

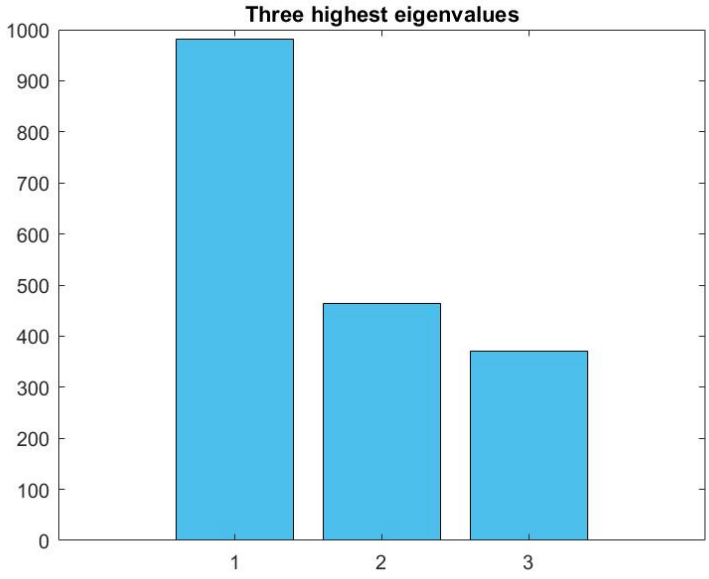


Figure 5.10: Plot of the highest three eigenvalues.

5.3.2. Alternative lists of candidates

Using the PCA, it is possible to crate new lists of candidates among the subjects. The lists made with the PCA gives the candidates who are more representative of the predominant personality trends. Lists are created selecting the candidates that maximize or minimizes the projection of the responses in the three principal components. Candidates who are in the extreme of the lists, are the ones who are more dissimilar in this Principal Component. Unlike the Big Five lists generated in the previous chapter, this classification doesn't provide a psychological description of the subjects. However, it provides the list of the candidates who are expected to provide the maximum information.

In Figure 5.11 the differences between both lists can be seen:

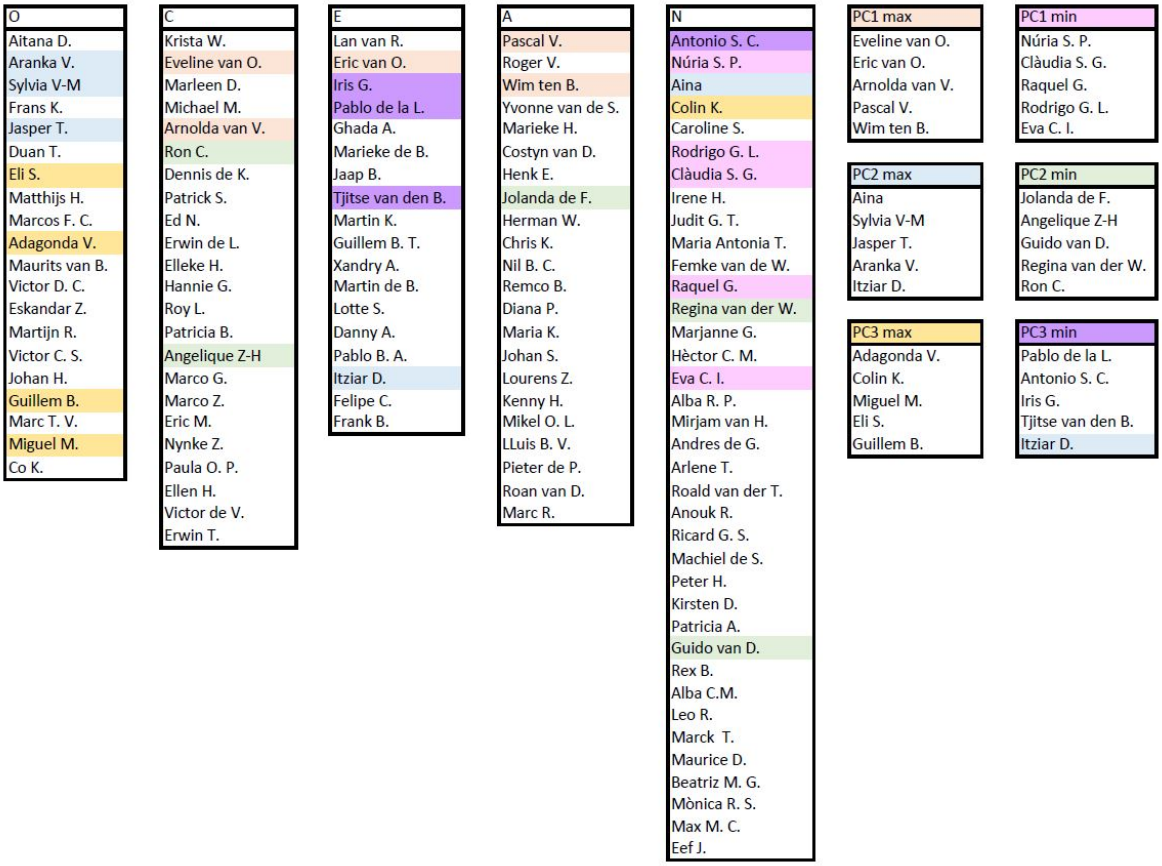


Figure 5.11: Lists with the Big Five Selection 2 and PCA.

Most of the participants who are in the top list of the PCA are also in the lists made by the BF. It can be seen that in the majority of the cases, people from the extremes of the different principal components have different personalities. For example, in PC1, people from the maximum extreme are in the top of the categories C, E and A, and in the minimum extreme there are only people from the category N.

The objective of this section is to obtain the candidates who stand out the most among all the data base.

Nonetheless, it seems that there is a correlation between the PCA and the Big Five, but in order to obtain more conclusions further investigation need to be done. Also it would be convenient to have much more data.

The Matlab algorithm to extract PCA components is attached on the appendix H.

6

Conclusions and Future work/Trends

It is well known that Automated Vehicles are a technology that sooner or later will become conventional in our every-day life. The automobile industry, including companies such Tesla, Nissan, Volvo and Volkswagen, is deeply involved investing in this technology. Automated vehicles not only provide a safer driving for experienced users but also a more comfortable and accessible experience for all. Human error is still the principal cause of accidents and this is where artificial intelligence might have an impact. Nonetheless, there is still much work that needs to be done for making automated vehicles a reality. In order to make this possible, many laws and regulations need to be changed and technology needs to consider all possible hazards.

TU Delft considered to develop an experiment to analyze human behavior when using Automated Vehicles. The goal is to consider the reactions of humans and to create confidence to the human user. As a consequence to develop better technology, specially for Human Machine Interfaces (HMI). This project proposes two different ways to make the selection of the candidates.

The first selection of significant drivers, based on The Big Five inventory, it is a classification of the personality of the participants based on five dimensions. It selects from a big database the best candidates representing each dimension: openness, conscientiousness, neuroticism, agreeableness and extraversion. This method of personality classification was chosen since it's simple to implement and it has been studied for many years by psychologists. The lists made by the Big Five inventory give insights regarding the personality of the candidates and thus a general trend on how their driving behaviour might be. It is important to emphasize that the results do not necessary mean that someone belongs only to his or her assigned category,

but only that they fit better than others in the data pool. Two programs have been developed in Matlab code to choose the best candidates for the experiments. Both are based on ordered lists.

On the other hand, Principal Component Analysis is an interesting alternative. It is a method which gives a classification of candidates who stand out in some features which are believed to contain the maximum information. This is a procedure that is useful in very different datasets corresponding to different applications. The figure representing the candidates along the two principal components indicates that there are no clusters, which means that the personalities of the participants are uniformly distributed. The PCA however, doesn't not indicate which psychological feature dominates since it is based purely on a mathematical analysis of the dataset. Nevertheless, it can be observed that some of the subjects are in the top of both lists, suggesting a possible relation between the two analysis techniques.

Both methods give consistent candidates and contain different kind of information. Depending on the purpose of the selection, one or the other might be better. If the psychological information about the category is important, selecting the candidates using the BF test might be better. But if the purpose is to select candidates which stand out from the rest, but without knowing their psychological profile, it would be convenient to use the PCA which provides, in theory, the maximum information possible.

As for the future work, it would be important to analyse the outcome of the experiment to be performed at TU Delft to see if there is a different pattern for each category. If no different pattern behaviours are identified using the Big Five Inventory, then using the PCA might be a promising alternative.

Bibliography

[1] Kooij, Julian (2018). Intelligent Vehicles (ME41105). Delft University of Technology. Delft, The Netherlands.

[2] Shapiro, Danny (2019). Leveling Up: What Is Level 2 Automated Driving? Advanced driver assistance systems help, but Level 2+ brings even greater vehicle safety capabilities. Nvidia. Retrieved from: https://blogs.nvidia.com/blog/2019/02/06/what-is-level-2-automated-driving/

[3] DGT (2017). 1160 fallecidos en accidente de tráfico. Retrieved from: http://revista.dgt.es/es/noticias/nacional/2017/01ENERO/0103balance-accidentes-2016.shtml.XYnu2ygzY2x

[4] Howard, Bill (2014). V2V: What are vehicle-to-vehicle communications and how do they work?. ExtremeTech. Retrieved from: https://www.extremetech.com/extreme/176093-v2v-what-are-vehicle-to-vehicle-communications-and-how-does-it-work

[5] M. Pinelis (2013). “Automotive sensors and electronics: Trends and developments in 2013”, in Automot. Sensors Electron. Expo, Detroit,MI, USA, Oct. 2013.

[6] N. Lu, N. Cheng, N. Zhang, X. Shen and J. W. Mark, (2014). "Connected Vehicles: Solutions and Challenges," in IEEE Internet of Things Journal, vol. 1, no. 4, pp. 289-299, Aug. 2014. Retrieved from: https://ieeexplore.ieee.org/abstract/document/6823640

[7] Othmane L.B., Weffers H., Mohamad M.M., Wolf M. (2015) A Survey of Security and Privacy in Connected Vehicles. In: Benhaddou D., Al-Fuqaha A. (eds) Wireless Sensor and Mobile Ad-Hoc Networks. Springer, New York, NY

[8] L. Jackson, P. Chapman, D. Crundall, (2019).“What happens next? Predicting other road users’ behaviour as a function of driving experience and processing time”. April 2019. Retrieved from: https://www.tandfonline.com/doi/abs/10.1080/00140130802030714

[9] Panagiotis Mitropoulos; Gerardo Cupido; and Manoj Namboodiri. (2009). "Cognitive Approach to Construction Safety: Task Demand-Capability Model". Volume 135 Issue 9 - September 2009. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000060

[10] Roberts, B. W., & Jackson, J. J. (2008). Sociogenomic personality psychology. Journal of personality, 76(6), 1523–1544. doi:10.1111/j.1467-6494.2008.00530.x

[11] Cherry, Kendra (2019). The 4 Major Personality Perspectives. Verywellmind. Approved by Wellness Board expert Amy Morin, LCSW. Retrieved from: https://www.verywellmind.com/personality-perspectives-2795950

[12] Cherry, Kendra (2019). The Big Five Personality Traits. 5 Major Factors of Personality. Verywellmind. Medically reviewed by Steven Gans, MD. Retrieved from: https://www.verywellmind.com/the-big-five-personality-dimensions-2795422

[13] Ruíz, José (2019). Modelo de los Big Five. "Psicología de la Peraonalidad". Universitat de Barcelona. Barcelona, Spain.

[14] Vital, Anna (2018). Big Five Personality Traits – Infographic. adioma. Retrieved from:https://blog.adioma.com/5-personality-traits-infographic/

[15] Cherry, Kendra (2019). Freud's Id, Ego, and Superego. Verywellmind. Medically reviewed by a board-certified physician. Retrieved from: https://www.verywellmind.com/the-id-ego-and-superego-2795951

[16] Boag, S (2014). Ego, drives, and the dynamics of internal objects. Front Psychol. 2014;5:666. doi:10.3389/fpsyg.2014.00666

[17] Cherry, Kendra (2019). Psychologist Hans Eysenck Biography. Verywellmind. Retrieved from: https://www.verywellmind.com/hans-eysenck-1916-1997-2795509

[18] McLeod, S. A. (2017). Theories of personality. Retrieved from https://www.simplypsychology.org/personality-theories.html

[19] Mahony, Patrick (2016). Eysenck's Personality Inventory (EPI) (Extroversion/Introversion). Patrick Mahony Psychoanalytic & Psychoanalysis. Retreived from: https://patrickmahony.co.uk/test/eysencks-personality-inventory-epi-extroversionintroversion/

[20] Cherry, Kendra (2019). Cattell's 16 Personality Factors. Verywellmind. Retrieved from: https://www.verywellmind.com/cattells-16-personality-factors-2795977

[21] Vital, Mark (2019). The 16 Personality Types by Myers-Briggs and Keirsey. adioma. Retrieved from: https://blog.adioma.com/16-personality-types/

[22] Slovic, P. (1987). "Perception of Risk." Science 236 (17 April): 280-285. Retrieved from: http://www.heatherlench.com/wp-content/uploads/2008/07/slovic.pdf

[23] Dick de Waard (1996). “The Measurement of Drivers’ Mental Workload”. Retrieved from: http://apps.usd.edu/coglab/schieber/pdf/deWaard-Thesis.pdf

[24] Fuller, R. (2005). "Towards a general theory of driver behaviour". Accid Anal Prev. 2005 May;37(3):461-72. DOI: 10.1016/j.aap.2004.11.003 PMID: 15784200.

[25] Philipp Yorck Herzberg, (2009). Beyond “accident-proneness”: Using Five-Factor Model prototypes to predict driving behavior, Journal of Research in Personality, Volume 43, Issue 6, 2009, Pages 1096-1100, ISSN 0092-6566, https://doi.org/10.1016/j.jrp.2009.08.008. (http://www.sciencedirect.com/science/article/pii/S0092656609001846)

[26] Bahram Esmaeili, Hamid Reza Imani Far, Hossein Hosseini, & Mohammad Sharifi. (2012). The Relationship between Personality Characteristics and Driving Behavior (Version 9000). http://doi.org/10.5281/zenodo.1071942

[27] Shen B, Qu W, Ge Y, Sun X, Zhang K (2018) The relationship between personalities and self-report positive driving behavior in a Chinese sample. PLoS ONE 13(1): e0190746. https://doi.org/10.1371/journal.pone.0190746

[28] Arthur, W., & Graziano, W. G. (1996). The five factor model, conscientiousness, and driving accident involvement. Journal of Personality, 64, 593–618.

[29] Arthur, W., Tubre, T., Day, E. A., Sheehan, M. K., Sanchez-Ku, M. L., Paul, D., et al. (2001). Motor vehicle crash involvement and moving violations: Convergence of self-report and archival data. Human Factors, 43, 1–11.

[30] Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Safety Science, 41, 427–443.

[31] Galarnyk, Michael (2017). PCA using Python (scikit-learn). Towards Data Science. Medium. Retrieved from: https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

[32] Starmer, Josh (2016). StatQuest: Linear Discriminant Analysis (LDA), clearly explained. PRINCIPAL COMPONENT ANALYSIS. STATQUEST An epic journey through computational methods for molecular geneticists. Retrieved from: https://statquest.org/tag/principal-component-analysis/

[33] Serrano, Luis (2019). Dimensionality Reduction (PCA). Machine Learning Fundamentals. Luis G. Serrano, Machine learning made easy. Retrieved from: https://serrano.academy/machine-learning/

A

Automated driving and the Big Five Personality Traits in English

Automated driving and the Big Five Personality Traits: One size fits none?

Do you want to be a participant in a driving simulator experiment that focuses on automated driving?

Here at TU Delft, we are preparing a project to investigate the behaviour of drivers of automated vehicles, based on the Big Five Personalities.

Our main goal with this questionnaire is to recruit participants of whom we think will be representative for our project. The experiment will be performed in our driving simulator, located here at TU Delft, and will in total take no longer than 1 hour of your time.

Please note that filling in this questionnaire does not automatically mean that you will be selected to take part in this experiment. You will be notified if you have been selected. If you are still interested, please fill in the fields below, and continue reading and answering the questions while scrolling down.

Please make sure you have read the consent form before answering and/or submitting the questions.

Questions with * are necessary.

- Email address*

_____

- Name and Surname*

_____

CONSENT FORM

Any information provided will be handled with the utmost care, and privacy is given priority at all times. None other than the researcher will handle this information, which in question, is henceforth considered to be responsible, and thereby acknowledges his or her responsibility.

All information collected by the researcher about yourself during the participation in this study will be safely secured all times, stored on a password protected computer. This information will only be used for the purpose of this study. All directly identifiable data, such as names and email addresses, that will have been collected will be deleted after the experiment has been completed.

Your participation is voluntary, which means you can withdraw at any given time, either during the completion of this questionnaire, or during the performance of the driving simulator experiment.

Please accept the terms and conditions if you agree with the statements:

- I have read and understood the information provided*

O Yes

O No

- I agree to take part in this project and agree for my data to be used for the purpose of this study*

O Yes

O No

- I understand my participation is voluntary*

O Yes

O No

- I understand I am not automatically selected after filling the questionnaire*

O Yes

O No

- Telephone number (This will only be used to contact you if needed):

_____

- In what city are you currently living in? *

_____

AVAILABILITY AND DEMOGRAPHIC QUESTIONNAIRE

- Will you be available this summer to take part in our driving simulator experiment? *

O Yes

O No

O Maybe

- Age*

_____ Years

- Gender*

O Female

O Male

O I'd rather not say

- Where are you from? *

_____

- Education *

O No school completion

O Primary School

O Secondary School

O Vocational School

O Higher Education – BSc/BA

O Higher Education – Msc/MA

O Higher Education – PhD

O Other

- Profession *

_____

- Do you use contact lenses/glasses? *

O Yes

O No

- Do you have a driving licence? *

O Yes

O No

- If the answer is yes, how many years of driving experience do you have?

_____ Years

- Average driving engagement in the last 12 months:

O Every day O 4-6 days a week O 1-3 days a week O Once a month

O Less than once a month O Never

- Average kilometres in the last 12 months *

O	0	O	1-1,000	O	1,001-5,000	O	5,001-10,000
O	10,001-20,000	O	20,001-30,000	O	30,001-50,000	O	50,000+

- Do you have any experience in automated driving systems?*

O Yes

O No

- If the answer is yes, how many years of experience in automated driving systems do you have?

_____ Years

- If the answer is yes, which one?

O ACC

O Lane keeping assistance

O Others _____

- Average kilometres in the last 12 months in automated driving *

O	0	O	1-1,000	O	1,001-5,000	O	5,001-10,000
O	10,001-20,000	O	20,001-30,000	O	30,001-50,000	O	50,000+

- Do you take any kind of drugs at present?*

O Yes

O No

O I'd rather not say

- If the answer is yes, what kind of drugs to you normally take and how often do you take them? *

	Every day	4-6 days a week	1-3 days a week	Once a month	Less than once a month	Never	I'd rather not say
Alcohol							
Cannabis							
Tobacco							
Cocaine							
Ecstasy							
LSD							
Other							

BIG FIVE QUESTIONNAIRE

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please select a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 Disagree Strongly	2 Disagree a little	3 Neither agree nor disagree	4 Agree a little	5 Agree strongly

I am someone who…

1. _____ Is talkative
2. _____ Tends to find fault with others
3. _____ Does a thorough job
4. _____ Is depressed, blue
5. _____ Is original, comes up with new ideas
6. _____ Is reserved
7. _____ Is helpful and unselfish with others
8. _____ Can be somewhat careless
9. _____ Is relaxed, handles stress well.
10. _____ Is curious about many different things
11. _____ Is full of energy
12. _____ Starts quarrels with others
13. _____ Is a reliable worker
14. _____ Can be tense
15. _____ Is ingenious, a deep thinker
16. _____ Generates a lot of enthusiasm
17. _____ Has a forgiving nature
18. _____ Tends to be disorganized

19. _____ Worries a lot

20. _____ Has an active imagination

21. _____ Tends to be quiet

22. _____ Is generally trusting

23. _____ Tends to be lazy

24. _____ Is emotionally stable, not easily upset

25. _____ Is inventive

26. _____ Has an assertive personality

27. _____ Can be cold and aloof

28. _____ Perseveres until the task is finished

29. _____ Can be moody

30. _____ Values artistic, aesthetic experiences

31. _____ Is sometimes shy, inhibited

32. _____ Is considerate and kind to almost everyone

33. _____ Does things efficiently

34. _____ Remains calm in tense situations

35. _____ Prefers work that is routine

36. _____ Is outgoing, sociable

37. _____ Is sometimes rude to others

38. _____ Makes plans and follows through with them

39. _____ Gets nervous easily

40. _____ Likes to reflect, play with ideas

41. _____ Has few artistic interests

42. _____ Likes to cooperate with others

43. _____ Is easily distracted

44. _____ Is sophisticated in art, music, or literature

B

Automated driving and the Big Five Personality Traits in Dutch

Geautomatiseerd rijden en de Big Five Personality Traits: Eén maat past niemand?

Wilt u deelnemen aan een rijsimulatorexperiment dat zich richt op geautomatiseerd rijden?

Bij de TU Delft bereiden we een project voor om het gedrag van bestuurders van (deels) zelfrijdende voertuigen te onderzoeken, gebaseerd op de Big Five Personalities.

Ons belangrijkste doel met deze vragenlijst is het werven van deelnemers waarvan wij denken dat ze representatief zijn voor ons project. Het experiment wordt uitgevoerd in de rijsimulator aan de TU Delft, en zal in totaal niet langer dan 1 uur van uw tijd in beslag nemen.

Houd er rekening mee dat het invullen van deze vragenlijst niet automatisch betekent dat u wordt geselecteerd om deel te nemen aan dit experiment. U krijgt een bericht als u geselecteerd bent. Als u nog steeds geïnteresseerd bent, vul dan de onderstaande velden in, en ga door met het lezen en beantwoorden van de vragen terwijl u naar beneden scrollt.

Zorg ervoor dat u het toestemmingsformulier heeft gelezen voordat u de vragen beantwoordt en/of indient.

Vragen met * zijn noodzakelijk.

- E-mail adres*

_____

- Naam en achternaam*

_____

TOESTEMMINGSFORMULIER

Alle verstrekte informatie zal met de grootst mogelijke zorgvuldigheid worden behandeld, waarbij privacy te allen tijde voorop staat. Niemand anders dan de onderzoeker zal deze informatie verwerken en inzien en erkent hierbij zijn/haar verantwoordelijkheid hiervoor.

Alle informatie die de onderzoeker tijdens de deelname aan dit onderzoek verzamelt, wordt te allen tijde veilig opgeslagen op een met een wachtwoord beveiligde computer. Deze informatie wordt alleen gebruikt voor het doel van dit onderzoek. Alle direct identificeerbare gegevens, zoals namen en e-mailadressen, die zijn verzameld, worden na afloop van het experiment verwijderd.

Uw deelname is vrijwillig, wat betekent dat u zich op elk moment kunt terugtrekken, zowel tijdens het invullen van deze vragenlijst, als tijdens het uitvoeren van het rijsimulatorexperiment.

Accepteer de algemene voorwaarden als u akkoord gaat met de verklaringen:

- Ik heb de verstrekte informatie gelezen en begrepen.*

O Ja

O Nee

- Ik ga akkoord met deelname aan dit project en ga ermee akkoord dat mijn gegevens worden gebruikt voor dit onderzoek.*

O Ja

O Nee

- Ik begrijp dat mijn deelname vrijwillig is.*

O Ja

O Nee

- Ik begrijp dat ik niet automatisch geselecteerd ben na het invullen van de vragenlijst.*

O Ja

O Nee

- Telefoonnummer: (dit wordt alleen gebruikt om contact met u op te nemen indien nodig):

_____

- Wat is uw woonplaats? *

_____

BESCHIKBAARHEID EN DEMOGRAFISCHE VRAGENLIJST

- Bent u deze zomer beschikbaar om deel te nemen aan ons rijsimulator-experiment? *

O Ja

O Nee

O Mogelijk

- Leeftijd*

_____ Jaar

- Geslacht*

O Vrouw

O Man

O Dit zeg ik liever niet.

- Wat is je woonplaats? *

_____

- Onderwijs*

O Geen schooldiploma

O Basisschool

O Middelbare school

O MBO

O HBO (BSc/BA)

O Universiteit (Msc/MA)

O Andere

- Beroep *

_____

- Gebruikt u contactlenzen of een bril? *

O Ja

O Nee

- Heeft u een B-rijbewijs? *

O Ja

O Nee

- Indien ja, hoeveel jaar rijervaring heeft u?

_____ Jaren

- Hoeveel dagen in de laatste 12 maanden bestuurde u (gemiddeld) een auto:

O Elke dag O 4-6 dagen per week O 1-3 dagen per week O Eenmaal per maand

O Minder dan één keer per maand O Nooit

- Hoeveel kilometer heeft u in de laatste 12 maanden gereden?

O	0	O	1-1,000	O	1,001-5,000	O	5,001-10,000
O	10,001-20,000	O	20,001-30,000	O	30,001-50,000	O	50,000+

- Heeft u enige ervaring met ADAS (ondersteunende rijsystemen)? *

O Ja

O Nee

- Indien ja, hoelang heeft u ervaring met ADAS (ondersteunende rijsystemen)

_____ Jaar

- Zo ja, welke ADAS gebruikt u?

O Adaptive cruise control (automatisch afstand van voorganger aanhouden)

O Lane Keeping Assist (Hulp bij het aanhouden van rijstroken)

O Andere _____

- Hoeveel kilometer heeft u in de laatste 12 maanden gereden met behulp van ADAS?

O	0	O	1-1,000	O	1,001-5,000	O	5,001-10,000
O	10,001-20,000	O	20,001-30,000	O	30,001-50,000	O	50,000+

- Neemt u op dit moment drugs? *

O Ja

O Nee

O Dit zeg ik liever niet.

- Indien ja, wat voor soort drugs gebruikt u en hoe vaak gebruikt u deze? *

	Elke dag	4-6 dagen per week	1-3 dagen per week	Eenmaal per maand	Minder dan één keer per maand	Nooit	Dit zeg ik liever niet
Alcohol							
Cannabis							
Tabak							
Cocaïne							
XTC							
LSD							
Andere							

GROTE VIJF VRAGENSTELLER

Hieronder staan een aantal persoonskenmerken beschreven die mogelijk op u van toepassing zijn. Geef iedere stelling een nummer om aan te geven in hoeverre u het eens of oneens bent met die stelling.

1 Compleet oneens	2 Beetje oneens	3 Eens noch oneens	4 Beetje eens	5 Compleet eens

Ik ben iemand die…

1. _____ Spraakzaam is

2. _____ Geneigd is kritiek te hebben op anderen

3. _____ Grondig te werk gaat

4. _____ Somber is

5. _____ Origineel is, met nieuwe ideeën komt

6. _____ Terughoudend is

7. _____ Behulpzaam en onzelfzuchtig ten opzichte van anderen is

8. _____ Een beetje nonchalant kan zijn

9. _____ Ontspannen is, goed met stress kan omgaan

10. _____ Benieuwd is naar veel verschillende dingen

11. _____ Vol energie is

12. _____ Snel ruzie maakt

13. _____ Een werker is waar men van op aan kan

14. _____ Gespannen kan zijn

15. _____ Scherpzinnig, een denker is

16. _____ Veel enthousiasme opwekt

17. _____ Vergevingsgezind is

18. _____ Doorgaans geneigd is tot slordigheid

19. _____ Zich veel zorgen maakt

20. _____ Een levendige fantasie heeft

21. _____ Doorgaans stil is

22. _____ Mensen over het algemeen vertrouwt

23. _____ Geneigd is lui te zijn

24. _____ Emotioneel stabiel is, niet gemakkelijk overstuur raakt

25. _____ Vindingrijk is

26. _____ Voor zichzelf opkomt

27. _____ Koud en afstandelijk kan zijn

28. _____ Volhoudt tot de taak af is

29. _____ Humeurig kan zijn

30. _____ Waarde hecht aan kunstzinnige en esthetische ervaringen

31. _____ Soms verlegen of geremd is

32. _____ Attent en aardig is voor bijna iedereen

33. _____ Dingen efficiënt doet

34. _____ Kalm blijft in gespannen situaties

35. _____ Een voorkeur heeft voor werk dat routine is

36. _____ Hartelijk, een gezelschapsmens is

37. _____ Soms grof tegen anderen is

38. _____ Plannen maakt en deze doorzet

39. _____ Gemakkelijk zenuwachtig wordt

40. _____ Graag nadenkt, met ideeën speelt

41. _____ Weinig interesse voor kunst heeft

42. _____ Graag samenwerkt met anderen

43. _____ Gemakkelijk afgeleid is

44. _____ Het fijne weet van kunst, muziek of literatuur

C

Big Five Inventory in English

How I am in general

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which **you agree or disagree with that statement.**

1 Disagree Strongly	**2** Disagree a little	**3** Neither agree nor disagree	**4** Agree a little	**5** Agree strongly

I am someone who…

1. _____ Is talkative
2. _____ Tends to find fault with others
3. _____ Does a thorough job
4. _____ Is depressed, blue
5. _____ Is original, comes up with new ideas
6. _____ Is reserved
7. _____ Is helpful and unselfish with others
8. _____ Can be somewhat careless
9. _____ Is relaxed, handles stress well.
10. _____ Is curious about many different things
11. _____ Is full of energy
12. _____ Starts quarrels with others
13. _____ Is a reliable worker
14. _____ Can be tense
15. _____ Is ingenious, a deep thinker
16. _____ Generates a lot of enthusiasm
17. _____ Has a forgiving nature
18. _____ Tends to be disorganized
19. _____ Worries a lot
20. _____ Has an active imagination
21. _____ Tends to be quiet
22. _____ Is generally trusting
23. _____ Tends to be lazy
24. _____ Is emotionally stable, not easily upset
25. _____ Is inventive
26. _____ Has an assertive personality
27. _____ Can be cold and aloof
28. _____ Perseveres until the task is finished
29. _____ Can be moody
30. _____ Values artistic, aesthetic experiences
31. _____ Is sometimes shy, inhibited
32. _____ Is considerate and kind to almost everyone
33. _____ Does things efficiently
34. _____ Remains calm in tense situations
35. _____ Prefers work that is routine
36. _____ Is outgoing, sociable
37. _____ Is sometimes rude to others
38. _____ Makes plans and follows through with them
39. _____ Gets nervous easily
40. _____ Likes to reflect, play with ideas
41. _____ Has few artistic interests
42. _____ Likes to cooperate with others
43. _____ Is easily distracted
44. _____ Is sophisticated in art, music, or literature

SCORING INSTRUCTIONS

To score the BFI, you'll first need to **reverse-score** all negatively-keyed items:

Extraversion: 6, 21, 31
Agreeableness: 2, 12, 27, 37
Conscientiousness: 8, 18, 23, 43
Neuroticism: 9, 24, 34
Openness: 35, 41

To recode these items, you should subtract your score for all reverse-scored items from 6. For example, if you gave yourself a 5, compute 6 minus 5 and your recoded score is 1. That is, a score of 1 becomes 5, 2 becomes 4, 3 remains 3, 4 becomes 2, and 5 becomes 1.

Next, you will create scale scores by ***averaging*** the following items for each B5 domain (where R indicates using the reverse-scored item).

Extraversion: 1, 6R 11, 16, 21R, 26, 31R, 36
Agreeableness: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42
Conscientiousness: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R
Neuroticism: 4, 9R, 14, 19, 24R, 29, 34R, 39
Openness: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

SPSS SYNTAX

***** REVERSED ITEMS**

```
RECODE
  bfi2 bfi6 bfi8 bfi9 bfi12 bfi18 bfi21 bfi23 bfi24 bfi27 bfi31 bfi34 bfi35
  bfi37 bfi41 bfi43
  (1=5)  (2=4)  (3=3)  (4=2)  (5=1)  INTO  bfi2r bfi6r bfi8r bfi9r bfi12r bfi18r bfi21r bfi23r bfi24r
  bfi27r bfi31r bfi34r bfi35r bfi37r bfi41r bfi43r.
EXECUTE .
```

***** SCALE SCORES**

```
COMPUTE bfie = mean(bfi1,bfi6r,bfi11,bfi16,bfi21r,bfi26,bfi31r,bfi36) .
VARIABLE LABELS bfie 'BFI Extraversion scale score.
EXECUTE .

COMPUTE bfia = mean(bfi2r,bfi7,bfi12r,bfi17,bfi22,bfi27r,bfi32,bfi37r,bfi42) .
VARIABLE LABELS bfia 'BFI Agreeableness scale score' .
EXECUTE .

COMPUTE bfic = mean(bfi3,bfi8r,bfi13,bfi18r,bfi23r,bfi28,bfi33,bfi38,bfi43r) .
VARIABLE LABELS bfic 'BFI Conscientiousness scale score' .
EXECUTE .

COMPUTE bfin = mean(bfi4,bfi9r,bfi14,bfi19,bfi24r,bfi29,bfi34r,bfi39) .
VARIABLE LABELS bfin 'BFI Neuroticism scale score' .
EXECUTE .

COMPUTE bfio = mean(bfi5,bfi10,bfi15,bfi20,bfi25,bfi30,bfi35r,bfi40,bfi41r,bfi44) .
VARIABLE LABELS bfio 'BFI Openness scale score' .
EXECUTE .
```

REFERENCE INFORMATION

The BFI should be cited with the original and a more accessible, recent reference:

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.

D

Big Five Inventory in Dutch

Hoe ik doorgaans ben

Hier zijn een aantal karakteristieken die wel of niet op jou kunnen slaan. Bijvoorbeeld, ben je het er mee eens dat je iemand bent die *graag tijd met anderen doorbrengt*? Schrijf een getal naast elke bewering om aan te geven hoe erg je het er mee **eens of oneens bent met die bewering**.

1 Compleet oneens	**2** Beetje oneens	**3** Eens noch oneens	**4** Beetje eens	**5** Compleet eens

Ik ben iemand die…

1. _____ Spraakzaam is
2. _____ Geneigd is kritiek te hebben op anderen
3. _____ Grondig te werk gaat
4. _____ Somber is
5. _____ Origineel is, met nieuwe ideeën komt
6. _____ Terughoudend is
7. _____ Behulpzaam en onzelfzuchtig ten opzichte van anderen is
8. _____ Een beetje nonchalant kan zijn
9. _____ Ontspannen is, goed met stress kan omgaan
10. _____ Benieuwd is naar veel verschillende dingen
11. _____ Vol energie is
12. _____ Snel ruzie maakt
13. _____ Een werker is waar men van op aan kan
14. _____ Gespannen kan zijn
15. _____ Scherpzinnig, een denker is
16. _____ Veel enthousiasme opwekt
17. _____ Vergevingsgezind is
18. _____ Doorgaans geneigd is tot slordigheid
19. _____ Zich veel zorgen maakt
20. _____ Een levendige fantasie heeft
21. _____ Doorgaans stil is
22. _____ Mensen over het algemeen vertrouwt
23. _____ Geneigd is lui te zijn
24. _____ Emotioneel stabiel is, niet gemakkelijk overstuur raakt
25. _____ Vindingrijk is
26. _____ Voor zichzelf opkomt
27. _____ Kou den afstandelijk kan zijn
28. _____ Volhoudt tot de taak af is
29. _____ Humeurig kan zijn
30. _____ Waarde hecht aan kunstzinnige ervaringen
31. _____ Soms verlegen, geremd is
32. _____ Attent en aardig is voor bijna iedereen
33. _____ Dingen efficiënt doet
34. _____ Kalm blijft in gespannen situaties
35. _____ Een voorkeur heeft voor werk dat routine is
36. _____ Hartelijk, een gezelschapsmens is
37. _____ Soms grof tegen anderen is
38. _____ Plannen maakt en deze doorzet
39. _____ Gemakkelijk zenuwachtig wordt
40. _____ Graag nadenkt, met ideeën speelt
41. _____ Weinig interesse voor kunst heeft
42. _____ Graag samenwerkt met anderen
43. _____ Gemakkelijk afgeleid is
44. _____ Het fijne weet van kunst, muziek, of literatuur

SCORING INSTRUCTIONS

To score the BFI, you'll first need to **reverse-score** all negatively-keyed items:

Extraversion: 6, 21, 31
Agreeableness: 2, 12, 27, 37
Conscientiousness: 8, 18, 23, 43
Neuroticism: 9, 24, 34
Openness: 35, 41

To recode these items, you should subtract your score for all reverse-scored items from 6. For example, if you gave yourself a 5, compute 6 minus 5 and your recoded score is 1. That is, a score of 1 becomes 5, 2 becomes 4, 3 remains 3, 4 becomes 2, and 5 becomes 1.

Next, you will create scale scores by ***averaging*** the following items for each B5 domain (where R indicates using the reverse-scored item).

Extraversion: 1, 6R 11, 16, 21R, 26, 31R, 36
Agreeableness: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42
Conscientiousness: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R
Neuroticism: 4, 9R, 14, 19, 24R, 29, 34R, 39
Openness: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

SPSS SYNTAX

***** REVERSED ITEMS**

```
RECODE
  bfi2 bfi6 bfi8 bfi9 bfi12 bfi18 bfi21 bfi23 bfi24 bfi27 bfi31 bfi34 bfi35
  bfi37 bfi41 bfi43
  (1=5)  (2=4)  (3=3)  (4=2)  (5=1)  INTO  bfi2r bfi6r bfi8r bfi9r bfi12r bfi18r bfi21r bfi23r bfi24r
  bfi27r bfi31r bfi34r bfi35r bfi37r bfi41r bfi43r.
EXECUTE .
```

***** SCALE SCORES**

```
COMPUTE bfie = mean(bfi1,bfi6r,bfi11,bfi16,bfi21r,bfi26,bfi31r,bfi36) .
VARIABLE LABELS bfie 'BFI Extraversion scale score.
EXECUTE .

COMPUTE bfia = mean(bfi2r,bfi7,bfi12r,bfi17,bfi22,bfi27r,bfi32,bfi37r,bfi42) .
VARIABLE LABELS bfia 'BFI Agreeableness scale score' .
EXECUTE .

COMPUTE bfic = mean(bfi3,bfi8r,bfi13,bfi18r,bfi23r,bfi28,bfi33,bfi38,bfi43r) .
VARIABLE LABELS bfic 'BFI Conscientiousness scale score' .
EXECUTE .

COMPUTE bfin = mean(bfi4,bfi9r,bfi14,bfi19,bfi24r,bfi29,bfi34r,bfi39) .
VARIABLE LABELS bfin 'BFI Neuroticism scale score' .
EXECUTE .

COMPUTE bfio = mean(bfi5,bfi10,bfi15,bfi20,bfi25,bfi30,bfi35r,bfi40,bfi41r,bfi44) .
VARIABLE LABELS bfio 'BFI Openness scale score' .
EXECUTE .
```

REFERENCE INFORMATION

The BFI should be cited with the original and a more accessible, recent reference:

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.

E

Big Five Invenroty in Spanish

Modelo de los Cinco Grandes rasgos de la personalidad

En el estudio de la psicología de la personalidad, el conocido como Modelo de los cinco grandes (en inglés, "Big Five") es un patrón en el estudio de la personalidad que examina la estructura de ésta a partir de cinco elementos amplios o rasgos de personalidad.

Los cinco grandes rasgos de personalidad, también llamados factores principales, suelen recibir los siguientes nombres: factor O (apertura a las nuevas experiencias), factor C (responsabilidad), factor E (extroversión), factor A (amabilidad) y factor N (neuroticismo o inestabilidad emocional).

* Obligatòria

1. **Adreça electrònica** *

2. **Nombre y Apellidos** *

3. **Género** *
   Maqueu només un oval.
   - Mujer
   - Hombre
   - Prefiero no decirlo
   - Altres:

4. **Edad** *

Cuestionario

Las respuestas de este cuestionario se utilizarán para poder hacer un estudio estadístico sobre el análisis de componentes principales. En otras palabras, pone en duda el clásico modelo de los Cinco Grandes rasgos de personalidad y busca otros patrones de relación con base matemática.

Para saber qué rasgo predomina más en tu personalidad debes contestar a las siguientes preguntas del 1 al 5 (teniendo en cuenta que 1 es que estás en total desacuerdo y 5 que estás completamente de acuerdo).

Todos los datos se analizarán de forma anónima.

5. **1. Eres hablador** *
   Maqueu només un oval.

1	2	3	4	5

6. **2. Tiendes a encontrar fallos en los demás** *
   Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

7. **3. Realizas los trabajos rigurosamente** *
   Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

8. **4. Eres una persona deprimida y triste** *
   Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

9. **5. Eres original, tienes nuevas ideas** *
   Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

10. **6. Eres reservado** *
    Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

11. **7. Eres útil y desinteresado con los demás** *
    Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

12. **8. Puedes ser algo descuidado** *
    Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

13. **9. Eres relajado, manejas bien el estrés** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

14. **10. Tienes curiosidad por muchas cosas diferentes** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

15. **11. Estás lleno de energía** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

16. **12. Empiezas discusiones con otros** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

17. **13. Eres un trabajador en el que se puede confiar** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

18. **14. Eres tenso** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

19. **15. Eres ingenioso, piensas en profundidad** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

20. **16. Generas mucho entusiasmo** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

21. **17. Tienes una naturaleza indulgente (perdona fácilmente)** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

22. **18. Tiendes a ser desorganizado** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

23. **19. Te preocupas mucho** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

24. **20. Tienes una imaginación activa** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

25. **21. Tiendes a ser callado** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

26. **22. Eres generalmente confiado** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

27. **23. Tiendes a ser perezoso** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

28. **24. Eres emocionalmente estable, no te enfadas fácilmente** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

29. **25. Tienes facilidad para inventar** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

30. **26. Expresas tu opinión de manera firme** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

31. **27. Puedes ser frío y distante** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

32. **28. Perseveras hasta que finalizas las tareas** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

33. **29. Tiendes a ser malhumorado** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

34. **30. Valoras experiencias artísticas, y estéticamente gratificantes** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

35. **31. A veces eres tímido, inhibido** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

36. **32. Eres considerado y amable** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

37. **33. Haces las cosas de manera eficiente** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

38. **34. Permaneces tranquilo en situaciones tensas** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

39. **35. Prefieres el trabajo rutinario** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

40. **36. Eres extrovertido, sociable** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

41. **37. A veces eres grosero con los demás** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

42. **38. Haces planes y los sigues** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

43. **39. Te pones nervioso fácilmente** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

44. **40. Te gusta reflexionar, jugar con ideas.** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

45. **41. Tienes pocos intereses artísticos** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

46. **42. Te gusta cooperar con otros** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

47. **43. Te distraes fácilmente** *
Maqueu només un oval.

1	2	3	4	5
◯	◯	◯	◯	◯

48. **44. Eres sofisticado en arte, música o literatura** *

Maqueu només un oval.

1	2	3	4	5
⬭	⬭	⬭	⬭	⬭

F

Selection algorithm 1 for Tu Delft selection, based on the Big Five Inventory

Selection algorithm 1 based on the Big Five Inventory

```
clear all;
clc;

%% Parameters

%You have to input the number of maxim participants that you want:
num_max = 10;

%% Coputations

%You have to input the cells where the names of the participants are:
[~,names]  = xlsread('Answers.xlsx','Hoja1','B1:CI1');

%You have to input the cells where the answers of the participants
are:
m = xlsread('Answers.xlsx','Hoja1');

s = length(names);

%Compute the score of each category:

E=[m(1,:)+6-m(6,:)+m(11,:)+m(16,:)+6-m(21,:)+m(26,:)+6-
m(31,:)+m(36,:)];
A=[6-m(2,:)+m(7,:)+6-m(12,:)+m(17,:)+m(22,:)+6-m(27,:)+m(32,:)+6-
m(37,:)+m(42,:)];
C=[m(3,:)+6-m(8,:)+m(13,:)+6-m(18,:)+6-
m(23,:)+m(28,:)+m(33,:)+m(38,:)+6-m(43,:)];
N=[m(4,:)+6-m(9,:)+m(14,:)+m(19,:)+6-m(24,:)+m(29,:)+6-
m(34,:)+m(39,:)];
O=[m(5,:)+m(10,:)+m(15,:)+m(20,:)+m(25,:)+m(30,:)+6-m(35,:)+m(40,:)+6-
m(41,:)+m(44,:)];

n=[E;A;C;N;O];

%We extract the mean and then divide by the standard deviation of each
category in
%order to have a matrix with all the normalized scores.

norm=(n-mean(n')')./std(n')';

[snor,index]=sort(norm','descend');

%Snor is the matrix with all the scores sorted from bigger to smaller.
%Index is the order where each person is in the sorted matrix.

%%
%In this section, you create a matrix with all the repeated
%volunteers: rep4.

rep=index(1:num_max,:);
rep3=zeros(5*num_max,1);
for j=1:num_max
    for i=1:5
        rep2=find(rep(:,:)==rep(j,i));
        if length(rep2)>1
            rep3(rep2)=rep(j,i);
        end
    end
end
```

```
rep4=reshape(rep3,[num_max,5]);
%%

%In this section, you create a matrix where the repeated volunteers
are
%evaluated, you keep the highest score and the lowest are replaced by
a -1: rep5.

rep5=rep4;

for j=1:num_max
    for i=1:5
        if rep5(j,i)>0
            for jj=1:num_max
                for ii=1:5
                    if rep5(jj,ii)==rep5(j,i)
                        if snor(j,i)<snor(jj,ii)
                            rep5(j,i)=-1;
                        end
                    end
                end
            end
        end
    end
end

rep5;
%%

%In this section, you replace the 0 for the no repeated volunteers.

rep6=rep5;

for j=1:num_max
    for i=1:5
        if rep6(j,i)==0
            rep6(j,i)=rep(j,i);
        end
    end
end


%%

%In this section, you create a matrix with the same number of new
volunteers as
%-1 we have in each column. Only adding the volunteers that aren't on
%the list, so there are no repetitions.

rep7=zeros(s,5);
rep7(1:num_max,1:5)=rep6;
for i=1:5
    add_element=sum(rep6(:,i)==-1); %We look for how many -1 there are
in each column.
    for kk=1:add_element
        kk2=kk;
        found_element=0;
        while (found_element==0)
            B=sum(sum(index(num_max+kk2,i)==rep7));
            if ((B)==0)
```

```
                rep7(num_max+kk,i)=index(num_max+kk2,i);
                found_element=1;
            else
                kk2=kk2+1;
            end
        end
    end
end



%%

%Finally, you create a matrix with the final volunteers without
repetitions.
%In addition, you create a matrix with the names of the most optimal
%volunteers for each category.

non_rep=zeros(num_max,5);

for i=1:5
    jj=1;
    j=1;
        while jj<(num_max+1)
            if rep7(j,i)~=-1
                non_rep(jj,i)=rep7(j,i);
                jj=jj+1;
            end
            j=j+1;
        end
end


non_rep;

%The final list:
disp('Category E list')
names(non_rep(:,1))'
disp('Category A list')
names(non_rep(:,2))'
disp('Category C list')
names(non_rep(:,3))'
disp('Category N list')
names(non_rep(:,4))'
disp('Category O list')
names(non_rep(:,5))'
```

G

Selection algorithm 2 for Tu Delft selection, based on the Big Five Inventory

Selection algorithm 2 based on the Big Five Inventory

```
clear all;
clc;


%% Computations


%You have to input the cells where the answers of the participants are:
m = xlsread('final_ans_names','Hoja1');

%You have to input the cells where the names of the participants are:
[~,names]  = xlsread('final_ans_names','Hoja1','B1:DQ1');

s = length(names);

%Compute the score of each category:

E=[m(1,:)+6-m(6,:)+m(11,:)+m(16,:)+6-m(21,:)+m(26,:)+6-m(31,:)+m(36,:)];
A=[6-m(2,:)+m(7,:)+6-m(12,:)+m(17,:)+m(22,:)+6-m(27,:)+m(32,:)+6-
m(37,:)+m(42,:)];
C=[m(3,:)+6-m(8,:)+m(13,:)+6-m(18,:)+6-m(23,:)+m(28,:)+m(33,:)+m(38,:)+6-
m(43,:)];
N=[m(4,:)+6-m(9,:)+m(14,:)+m(19,:)+6-m(24,:)+m(29,:)+6-m(34,:)+m(39,:)];
O=[m(5,:)+m(10,:)+m(15,:)+m(20,:)+m(25,:)+m(30,:)+6-m(35,:)+m(40,:)+6-
m(41,:)+m(44,:)];

n=[E;A;C;N;O];

%We extract the mean and then divide by the standard deviation of each
category in
%order to have a matrix with all the normalized scores.

norm=(n-mean(n')')./std(n')';

m_norm=norm.*(norm==max(norm))-100*(norm~=max(norm));

[snor,index]=sort(m_norm','descend');

%Snor is the matrix with all the scores sorted from bigger to smaller.
%Index is the order where each person is in the sorted matrix.

l_len=sum(snor~=-100);

disp('Category E list')
names(index(1:l_len(1),1))'
disp('Category A list')
names(index(1:l_len(2),2))'
disp('Category C list')
names(index(1:l_len(3),3))'
disp('Category N list')
names(index(1:l_len(4),4))'
disp('Category O list')
names(index(1:l_len(5),5))'
```

H

Selection algorithm of the PCA

Selection algorithm made with PCA

```
clear all;
clc;


%You have to input the cells where the answers of the participants are:
m = xlsread('final_ans_names','Hoja1');

%You have to input the cells where the names of the participants are:
[~,names]  = xlsread('final_ans_names','Hoja1','B1:DQ1');

s = length(names);

data = m';

[P,Q]=size(data); % P: number of people questioned, Q: number of
questions;

mdata=mean(data); % We compute the mean value of each question
mdata=ones(P,1)*mdata; % Transform mdata as a PxQ matrix in which all
elements of each columns contain the mean of the
% data matrix at that column
data2=data-mdata; % Now the data matrix contains the data in which the
collumns all have zero empirical mean

%Matrix with 0 mean.

%%

% Principal component analysis

L=5; % We decide how many of the most important components we want to
presenve in our data

[U, S, V]=svd(data2); % We compute the SVD of (zero mean) the data
matrix.

%It is a simple mode to obtain the values of eigenvectors and eigenvalues
from the covariance matrix.

% S^2= eigenvalues
% U= left singular vectors
% V= eigenvectors

SL=S(1:L,1:L); % We now consider only the L largest singular values
(=squareroot of the eigenvalues of the covariance)
UL=U(:,1:L); % and their corresponding left singular vectors
VL=V(:,1:L);% The VL matrix represents in each of the L columns the
combination of the questions that determine the
% L more significant features

% now we have a list of the same original people with the corresponding
score corresponding to the L principal components
% of the questionnaire. Both p_data contain the same information.
p_data3=data2*VL;
p_data2=UL*SL;
```

```
T_power=sum(diag(S'*S));
L_power=sum(diag(SL*SL));

%This is the amount of information that we get when computing with L PC.
Perc_power=L_power/T_power


%% Computations


[snor,index]=sort(p_data3,'descend');

%insert the top number of people
top=5


disp('PC1 max')
names(index(1:top,1))'
disp('PC1 min')
names(index(P:-1:P-top+1,1))'
disp('PC2 max')
names(index(1:top,2))'
disp('PC2 min')
names(index(P:-1:P-top+1,2))'
disp('PC3 max')
names(index(1:top,3))'
disp('PC3 min')
names(index(P:-1:P-top+1,3))'
```